Score: 2386.00   Matches: 460
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9   Indels: 0
DB: Gaps: 0

US-10-001-254-16 (1-460) x AK000528 (1-2820)

```
Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30 ATGAACAAACCCATAACACCATCAACACATTATGTGCGCTGCCTCAATGTTGGACTTATAGG  89

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   90 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 149

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  150 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGATTTGAAGCATTACT 209

Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr  80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  210 CAAACTGGAAAAAGTCCCACTTCTGAAATTACTGTTTGACTGGGGCACCACAAATTGCACA 269

Qy   81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  270 GTTGGTGATCTTGTTGATCTTTTGATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTG 329

Qy  101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLeuValAspLeuThrVal 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  330 CTCCCAGATGCTGTTCCCAAAACTGCTAATACACTTCCTTCATTAGTAGATCTAACAGTT 389

Qy  121 GlnGlnLysGlnMetProPheCysSerAspLysAspArgThrLeuMetThrProValGln 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  390 CAGCAAAAACAGATGCCTTTCTGTAGCGACAAAGACCACACTGTCCAGAT 449

Qy  141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  450 CTTGAACAAAGCTATATGCCACCTGACTCCTCAAGTCCAGAAAATAAAAGTTTAGAAGTT 509

Qy  161 SerAspThrArgPheHisSerPheTyrGluLeuLysAsnValThrAsnAsnPhe 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  510 AGTGATACACGTTTTCATTCATTTTATGAAATTGAAGAATGTCACAAATAACTT 569

Qy  181 AspGluArgProIleSerValGlyLysAsnLysMetGlyGlyGlyPheGlyValVal 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  570 GATGAAGACGACCCCATTTCTGTTGGTAATAAAATGGGAGGGGGATTTGGAGTTGTA 629

Qy  201 TyrArgSerGlyTyrValAsnAsnThrThrValAlaValLysLysIleLysLeuAsp 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  630 TATAGAAGCGGCTACGTAAATAACACAACAGTGGCAGTGAAGAAGCTTAAACTTGAC 689

Qy  221 IleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysValMetAlaLysCys 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  690 ATTACTACTGAAGAACTGAAACAGCAGTTTGATCAAGAAATAAAAGTAATGGCAAAGTGT 749

Qy  241 GlnHisGluAsnLeuValGlyLeuLeuGlyPheSerSerAspGlyAspLeuCysLeu 260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  750 CAACATGAAAACCTTAGTAGGACTACTTGGTTCTCCAAGTGATGGAGACCTCTGCTTA 809

Qy  261 ValThrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr 280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  810 GTTATAGTTTACATGCCTAATGGTTCATTGCTAGACAGACTCTCTTGCTTGGATGGTACT 869

Qy  281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAsnGlyIleAsn 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  870 CCACCACTTTCTTGGCACATGAGATGCAAGATTGCTCAGGGTGCAGCTAATGGCATCAAT 929

Qy  301 PheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp 320
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  930 TTTCTACATGAAAATCATCATATTCATAGAGATATAAAAAGTGCAAATATCTTACTGGAT 989

Qy  321 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla 340
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db
```

RESULT 3

AK000528

LOCUS AK000528 2820 bp mRNA linear PRI 22-FEB-2000
DEFINITION Homo sapiens cDNA FLJ20521 fis, clone KAT10395.
ACCESSION AK000528
VERSION AK000528.1 GI:7020683
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2820)
 AUTHORS Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Isogai,T.,
  Nakamura,Y., Isogai,T. and Sugano,S.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2820)
 AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T.,
  Shibahara,T., Tanaka,T. and Nakamura,Y.
 TITLE Direct Submission
 JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
  University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1,
  Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,
  Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
  International Trade and Industry of Japan; cDNA full insert
  sequencing; Research Association for Biotechnology; cDNA library
  construction, 5'- & 3'-end one pass sequencing; Departent of
  Virology and Human Genome Center, Institute of Medical Science,
  University of Tokyo (partly supported by Science and Technology
  Agency)
FEATURES Location/Qualifiers
 source 1..2820
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="KAT10395"
  /cell_line="KATO III"
  /cell_type="signet-ring cell carcinoma"
  /clone_lib="KAT"
  /note="cloning vector pME18SFL3"
 CDS 30..1412
  /codon_start=1
  /protein_id="BAA91232.1"
  /db_xref="GI:7020684"
  /translation="MNKPIPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
  DDRYNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
  DAVPKTANTLPSKEAITVQQKQMPFCSDKDRTLMTPVQNLEQSYMPPDSSSPENKSLEV
  SDTRFHSFFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
  VDITEELKQQFDQEIKWMAKCQHENHVLSLLLGDDLCLVVVMPNGSLDRLSC
  LDGTPPLSWMMRCKIAQGANGINFLHENNHHIHRDIKSANILLDEAFTAKISDFGLAR
  ASEKFAQTVWTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
  QQLLQEMTAS"
BASE COUNT 940 a 534 c 579 g 767 t
ORIGIN

(handwritten annotations: "DNA encoding Seq ID #16"; "AK000528" circled; "2820 bp" circled; "PRI 22-FEB-2000" circled; "Watanabe" and "(sites)" circled)

```
Qy  401 GluGluIleGluAspProGluGluLysThrIleGluGlnAspTyrIleAspLysMetAsnAsp 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1271 GAAGAAATTGAAGATCCAGAAGAAAAGACAATTGAAGATTATATTGATAAAATGAATGAT 1330

Qy  421 AlaAspSerThrSerValGlyValAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys 440
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1331 GCTGATTCCACTTCAGTTGGAGCTATGTACTCTGTTGCTAGTCAATGTCTTCGATGAAAAG 1390

Qy  441 LysAsnLysLysArgProAspIleLysValGlnHisGluLysGluMetThrAlaSer 460
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1391 AAAAATAAGAAACGTAAGATCAAACAGGTCAAGAGAGATGACAGCTTCT 1450
```

Alignment Scores:
Pred. No.: 1.26e-197 Length: 2820

Alignment pred. No

DB: US-10-001-254-16 (1-460) x AX196260 (1-1383)   6   Gaps: 0

RESULT 4

LOCUS   AX196260   1383 bp   DNA   linear   PAT 28-AUG-2001
DEFINITION   Sequence 2 from Patent WO0151641.
ACCESSION   AX196260
VERSION   AX196260.1 GI:15386462
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Wesche,H. and Li,S.
  TITLE   Irak-4: compositions and methods of use
  JOURNAL   Patent: WO 0151641-A 2 19-JUL-2001;
             Tularik Inc. (US)
FEATURES   Location/Qualifiers
     source   1..1383
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="human IL-1 receptor-associated kinase 4 (IRAK-4)
                     cDNA"
     CDS   1..1383
                     /note="human IRAK-4"
                     /codon_start=1
                     /protein_id="CAC50090.1"
                     /db_xref="GI:15386463"
                     /translation="MNKPITPSTYVRCLNVGLIIRKLSDFIDPQEGWMKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTAGDLVDLLIQNEFPAPASLLLP
                     DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTFVONLEQSYMPDDSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTELKQQFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPLSWHMRCKIAQGAANGINFLHENHHIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAVMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLLDIKEEIEDEEKIIEDIDKKMNDADSTSVEAMYSVASQCLHEKRNRPDIKKV
                     QQLLQEMTAS"

BASE COUNT   463 a   244 c   283 g   393 t
ORIGIN

Db   990 GAAGCTTTTACTTGCTAAAATATCTGACTTTGCCTTGACCCGGGGCTTCTGAGAAGTTTGCC 1049
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 60

Db 341 GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleIleArg 20

Db 1050 CAGACAGTCATCACTAGCAGAAATTGTGGGAACAACAGCTTATATGGCACCAGAAGCTTTG 1109
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGATTTGAAGCATTACTT 120

Db 361 ArgGlyIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle 380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys 40

Db 1110 CGTGGGAGAAATAACACCACCAAAATCTGATATTTACAGCTTTGGTGTGGTTTTACTAGAAATA 1169
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 121 AAGCTCTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCCGTTGCTATTAAA 180

Db 381 IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAspIleLys 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60

Db 1170 ATAACTGGACTTCCAGCTGTGGATGAACACCGTGAACCTCAGTTATTGCTAGATATTAAA 1229
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 181 CAAACTGGAAAGAGTCCCACTTCAGAATTACTGTTTGACTGGGGCACCACAAATTGCACA 240

Db 401 GluGluIleIleGluAspIleGluLysLysMetAsnAsp 420
          |||||||||||||||||||||||||||||||||||||||
Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80

Db 1230 GAAGAAATTGAAGATATAGAAAAGAAGATGAATGAT 1289
          ||||||||||||||||||||||||||||||||||||
Qy 241 GCTGGAGATCTTGTGGATCTTTTGATACAATTTTTTGCCCCTGCGAGTCTTTTG 300

Db 421 AlaAspSerThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys 440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPheProAlaSerLeuLeu 100

Db 1290 GCTGATTCCACTTCAGTTGAAGCTATGTACTCTGTTGCTAGTCAATGTCTCGCATGAAAAG 1349
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 301 CTCCCACAGATGCTGTTCCCAAAACTGCTAATACACTACCTTCTAAAGGAAGCTATAACAGTT 360

Db 441 LysAsnLysArgProAspIleLysLysValGlnGlnLeuLeuGlnGluMetThrAlaSer 460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120

Db 1350 AAAAATAAGAGACCAGACATTAAGAAGGTTCAACAGCTGCTTCAAGAGATGACAGCTTCT 1409
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 121 GlnGlnLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn 140

Qy 141 LeuGluGlnSerTyrMetProAspAspSerSerProGluAsnLysSerLeuGluVal 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 361 CAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGGACACTGTGACAGAAT 420

Qy 161 SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe 180
Qy 421 CTTGAACAAAGCTATATGCCAGACTCCTCCAAGTCCAGAAAATAAAAGTTTAGAAGTT 480

Qy 181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal 200
Qy 481 AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGAAGAATGTCACAAATAACTTT 540

Qy 201 TyrLysGlyTyrValAsnAsnThrThrValAlaValLysLysLeuAlaAlaMetValAsp 220
Qy 541 GATGAACGACCCATTTCTGTTGGTGGTAATAAAATGGGAGAAGGAGGATTTGGAGTTGTA 600

Qy 221 IleThrThrGluLeuLysGlnGlnPheAspGlnGluIleLysValMetAlaLysCys 240
Qy 601 TATAAAGGCTATGTTAATAATACAACAGTTGCAGTGAAGAAGCTTGCAGCAATGGTTGAC 660

Qy 241 GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu 260
Qy 661 ATTACTACTGAAGAACTGAAACAGCAGTTTGATCAAGAAATAAAAGTAATGGCAAAGTGT 720

Qy 261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr 280
Qy 721 CAACACGAAAAACTTAGTAGAACTACTTGGTTTTCCAAGTGATGGAGATGACCCTCGCTTA 780

Qy 281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn 300
Qy 781 GTATAGTTTTCATCACGCTAAGGTTCAATTGCTCAGACAGACTCTCTTGGATGGTACT 840

Qy 301 PheLeuHisGluAsnHisHisIleIleHisSerAlaAsnIleLeuLeuLeuAsp 320
Qy 841 CCACCACTTTCTTGGCACATGAGATGCCAAGAGTTGCTCAATGGCATCAAT 900

Qy 321 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla 340
Qy 901 TTTCTACATGAAAAATCATCAATTATTCATAGAGATTAAAAGTGCAAAATATCTTACTGGAT 960

Qy 341 GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu 360
Qy 961 GAAGCTTTTACTGCTAAAAATATCTGACTTTGGCCCTTGCACGGGCTTCTGAGAAGTTTGCC 1020

Alignment Scores:
Pred. No.:   1.11e-197
Score: 2382.00
Percent Similarity: 99.78%
Best Local Similarity: 99.83%
Query Match:

                          Length: 1383
                          Matches: 459
                          Conservative: 0
                          Mismatches: 1
                          Indels: 0

DNA of sea ID#5

RESULT 6

LOCUS   AK000528

DEFINITION   Homo sapiens cDNA FLJ20521 fis, clone KAT10395.
ACCESSION   AK000528
VERSION   AK000528.1 GI:7020683
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
               Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
               Nakamura,Y., Isogai,T. and Sugano,S.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2820)
AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
               Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE   Direct Submission
JOURNAL   Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
               University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1,
               Minato-ku Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp,
               Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
               International Trade and Industry of Japan; cDNA full insert
               sequencing project supported by Research Association for Biotechnology; cDNA library
               construction, 5'- & 3'-end one pass sequencing: Department of
               Virology and Human Genome Center, Institute of Medical Science,
               University of Tokyo (partly supported by Science and Technology
               Agency).

FEATURES   Location/Qualifiers
     source   1..2820
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="KAT10395"
                     /cell_line="KATO III"
                     /cell_type="signet-ring cell carcinoma"
                     /clone_lib="KAT"
                     /note="cloning vector pME18SFL3"
     CDS   30..1412
                     /codon_start=1
                     /protein_id="BAA91232.1"
                     /db_xref="GI:7020684"
                     /note="unnamed protein product"
                     /translation="MNKPIPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRREFEALLQTGKSPTSELLFDWGTTNCTAGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQKOMPFCDKDRTLMTPVQNLEQSYMPFDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGVVNNTVAVKKLAAM
                     VDITTEELKQOFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPPLSWHRCKIAQGAANGINFLHENHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     POLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                     QQLLQEMTAS"

BASE COUNT   940 a   534 c   579 g   767 t
ORIGIN

Qy	1 ACATAATCTGCGCTGCCTCCAAATGTTGGGACTAATTAAATTAGGAAGCTCTCAGATTTTTATTGAATCCT 60
Db	54 ACATAATGTGCGCTGCCTCCAAATGTTGGGACTAATTAGGAAGCTGTCAGATTTTTATTGAATCCT 113
Qy	61 CAAGAAGGATGGAAGAAGAGTTAGCTGTGTAGCTATTAAAAAAAACCATTCTGGTGATTGATGATAGATAC 120
Db	114 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATTCTGGTGATTGATGATAGATAC 173
Qy	121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCACTCAAACTGGAAAAAGTCCCACTTCT 180
Db	174 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAAACTGGAAAAAGTCCCACTTCT 225
Qy	181 GAATTACTGTTTGACTGGGGACCACCAACAAAATTGCACAGTTGGTGATCTTGTTGGATCTTTTG 240
Db	234 GAATTACTGTTTGACTGGGGCACCACCAACAAAATTGCACAGTTGGTGATCTTGTTGGATCTTTTG 293
Qy	241 ATCCAAAATGAAATTTTTTTGCTCCCTGCGAGTCTTTTGGCTCCCGAGATGCTGTTCCC 294
Db	294 ATCCAAAATGAAATTTTTTTGCTCCCTGCGAGTCTTTTGGCTCCCGAGATGCTGTTCCC 347

Query Match   100.0%; Score 294; DB 9; Length 2820;
Best Local Similarity   100.0%;
Matches 294; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

RESULT 7

LOCUS   AX196260   1383 bp   DNA   linear   PAT 28-AUG-2001

DEFINITION   Sequence 2 from Patent WO0151641.
ACCESSION   AX196260
VERSION   AX196260.1 GI:15386462
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Wesche,H. and Li,S.
TITLE   Irak-4: compositions and methods of use
JOURNAL   Patent: WO 0151641-A 2 19-JUL-2001;
               Tularik Inc. (US)

FEATURES   Location/Qualifiers
     source   1..1383
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="human IL-1 receptor-associated kinase 4 (IRAK-4)
                     cDNA"
     CDS   1..1383
                     /note="human IRAK-4"
                     /codon_start=1
                     /protein_id="CAC60090.1"
                     /db_xref="GI:15386463"
                     /translation="MNKPIPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRREFEALLQTGKSPTSELLFDWGTTNCTAGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQKOMPFCDKDRTLMTPVQNLEQSYMPFDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGVVNNTVAVKKLAAM
                     VDITTEELKQOFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPPLSWHRCKIAQGAANGINFLHENHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     POLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                     QQLLQEMTAS"

BASE COUNT   463 a   244 c   283 g   393 t
ORIGIN

Qy	1 ACATAATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTTATTGAATCCT 60
Db	25 ACATAATGTGCGCCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTCATCCT 84
Qy	61 CAAGAAGGATGGAAGAAGAGTTAGCTGTAGCTATTAAAAAAAACCATTCTGGTCGATGATGATAGATAC 120
Db	85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATGATAGATAC 144
Qy	121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAAACTGGAAAAAGTCCCACTTCT 180
Db	145 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAAACTGGAAAAAGTCCCACTTCT 204
Qy	181 GAATTACTGTTTGACTGGGGCACCACCAACAAAATTGCACAGTTGGTGATCTTGTTGGATCTTTTG 240
Db	205 GAATTACTGCTTTGACTGGGGCACCACCAACAAATTGCACAGCTGGTGATCTTGTTGGATCTTTTG 264
Qy	241 ATCCAAAATGAAATTTTTTTGCTCCCTGCGAGTCTTTTGGCTCCCGAGATGCTGTTCCC 294

Query Match   99.5%; Score 292.4; DB 6; Length 1383;
Best Local Similarity   99.7%;
Matches 293; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
CDS              /organism="Homo sapiens"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:9606"
50. .1432
                 /codon_start=1
                 /protein_id="CAD37285.1"
                 /note="unnamed protein product"
                 /db_xref="GI:21656188"
                 /translation="MNKPIPPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                 DDRVNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                 DAVPKTANTLPSKEAITVQOKQMPFCDKDRTLMTPVQNLEQSYMPPDDSSSPENKSLEV
                 SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGVVNNTVAVKLAAM
                 VDITEELKQOFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVIVYMPNGSLLDRLSC
                 LDGTPPLSWMMRCKIAOGAGAGNGINFLHENHIHRDIKSANILLDEAFTAKISDFGLAR
                 ASEKFAQTVWTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                 PQLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSGASQCRHEKKNKSPDIKKV
                 HQLLQEMTAS"

BASE COUNT     912 a      547 c      586 g      772 t
ORIGIN

Alignment Scores:
Pred. No.:              1.71e-60
Score:                  517.00        Length:             2817
Percent Similarity:     100.00%       Matches:               98
Percent Identity:       100.00%       Conservative:           0
Best Local Similarity:  100.00%       Mismatches:             0
Query Match:            6             Indels:                 0
DB:                                   Gaps:                   0

US-10-001-254-6 (1-98) x AX431318 (1-2817)

Qy      1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     74 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 133

Qy     21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 193

Qy     41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCAACTTCT 253

Qy     61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 313

Qy     81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 367

RESULT 6
AF155118
LOCUS          AF155118
DEFINITION     Homo sapiens putative protein kinase NY-REN-64 antigen mRNA,
               complete cds.
ACCESSION      AF155118
VERSION        AF155118.1  GI:5360130
KEYWORDS       .
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 2817)
AUTHORS        Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
               Jongeneel,V., Gure,A.O., Jager,D., Jager,E., Knuth,A., Chen,Y.-T.
               and Old,L.J.
TITLE          Antigens recognized by autologous antibody in patients with
               renal-cell carcinoma
JOURNAL        Int. J. Cancer 83 (4), 456-464 (1999)
MEDLINE        99438124
PUBMED         10508479
REFERENCE      2  (bases 1 to 2817)
AUTHORS        Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
```

```
                                                                    Jongeneel,V., Gure,A.O., Jager,D., Jager,E., Knuth,A., Chen,Y.-T.
                                                                    and Old,L.J.
TITLE          Direct Submission
JOURNAL        Submitted (28-MAY-1999) Ludwig Institute, Sloan-Kettering
               Institute, 1275 York Ave, New York, NY 10021, USA
FEATURES       Location/Qualifiers
  source       1..2817
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /cell_line="1973/10-4"
                 /cell_type="renal cell carcinoma"
  CDS          50. .1432
                 /codon_start=1
                 /product="putative protein kinase NY-REN-64 antigen"
                 /protein_id="AAD42884.1"
                 /db_xref="GI:5360131"
                 /translation="MNKPIPPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                 DDRVNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                 DAVPKTANTLPSKEAITVQOKQMPFCDKDRTLMTPVQNLEQSYMPPDDSSSPENKSLEV
                 SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGVVNNTVAVKLAAM
                 VDITEELKQOFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVIVYMPNGSLLDRLSC
                 LDGTPPLSWHMRCKIAOGAGAGNGINFLHENHIHRDIKSANILLDEAFTAKISDFGLAR
                 ASEKFAQTVWTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                 PQLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSGASQCRHEKKNKSPDIKKV
                 HQLLQEMTAS"

BASE COUNT     912 a      547 c      586 g      772 t
ORIGIN

Alignment Scores:
Pred. No.:              1.71e-60
Score:                  517.00        Length:             2817
Percent Similarity:     100.00%       Matches:               98
Percent Identity:       100.00%       Conservative:           0
Best Local Similarity:  100.00%       Mismatches:             0
Query Match:            9             Indels:                 0
DB:                                   Gaps:                   0

US-10-001-254-6 (1-98) x AF155118 (1-2817)

Qy      1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     74 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 133

Qy     21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 193

Qy     41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCAACTTCT 253

Qy     61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 313

Qy     81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 367

RESULT 7
AK000528
LOCUS          AK000528
DEFINITION     Homo sapiens cDNA FLJ20521 fis, clone KAT10395.
ACCESSION      AK000528
VERSION        AK000528.1  GI:7020683
KEYWORDS       oligo capping; fis (full insert sequence).
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (sites)
AUTHORS        Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
               Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
```

TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 2820)
AUTHORS   Nakamura,Y., Isogai,T. and Sugano,S.
AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
          Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE   Direct Submission
JOURNAL   Submitted (15-FEB-2000) Sunio Sugano, Institute of Medical Science,
          University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1,
          Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp,
          Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
          International Trade and Industry of Japan; cDNA full insert
          sequencing: Research Association for Biotechnology; cDNA library
          construction, 5'- & 3'-end one pass sequencing: Department of
          Virology and Human Genome Center, Institute of Medical Science,
          University of Tokyo (partly supported by Science and Technology
          Agency).

FEATURES   Location/Qualifiers
     source   1..2820
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="KAT10395"
                    /cell_line="KATO III"
                    /cell_type="signet-ring cell carcinoma"
                    /clone_lib="KAT"
                    /note="cloning vector pME18SFL3".
     CDS   30..1412
                    /note="unnamed protein product"
                    /codon_start=1
                    /db_xref="GI:7020684"
                    /protein_id="BAA91232.1"
                    /translation="MNKPITTPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                    DDRYNQFHIRRFEALLQTGKSPTSELFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                    DAVPKTANTLPSKEAITVQQKOMPFCDKDRTLMTPVONLEQSYMPPDSSSPENKSLEV
                    SDTRFH6FSFYELKNVTNNFDERPISVGGNKMGEGGFQGVVYKGVVNNTTVAVKKLAAM
                    VDITTEELKQOFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVVMPNGSLLDRLSC
                    LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                    ASEKPAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEITIGLPAVDEHRE
                    POLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                    QQLLQEMTAS"
BASE COUNT   940 a   534 c   579 g   767 t
ORIGIN

US-10-001-254-6 (1-98) x AK000528 (1-2820)

Alignment Scores:
Pred. No.:
Score: 517.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 1.71e-60

Length: 2820
Matches: 98
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro   20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ACATATGTGCGCTGCCTCAAGGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 113

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr   40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 173

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 233

Qy   61 GluLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu   80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 234 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGGATCTTGTGGATCTTG   293

Qy   81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro   98
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 294 ATCCAAAAATGAATTTTTTGCTCCTGCGGAGTCTTTTGCTCCCAGATGCTGTTCCC   347

RESULT 8

LOCUS   AX196260   1383 bp   DNA   linear   PAT 28-AUG-2001
DEFINITION Sequence 2 from Patent WO0151641.
ACCESSION   AX196260
VERSION   AX196260.1 GI:15386462
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Wesche,H. and Li,S.
TITLE   Irak-4: compositions and methods of use
JOURNAL   Patent: WO 0151641-A 2 19-JUL-2001;
            Tularik Inc. (US)

FEATURES   Location/Qualifiers
     source   1..1383
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /note="human IL-1 receptor-associated kinase 4 (IRAK-4)
                    cDNA"
     CDS   1..1383
                    /note="human IRAK-4"
                    /codon_start=1
                    /protein_id="CAC60090.1"
                    /db_xref="GI:15386463"
                    /translation="MNKPITTSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                    DDRYNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                    DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVONLEQSYMPPDSSSPENKSLEV
                    SDTRFH6FSFYELKNVTNNFDERPISVGGNKMGEGGFQGVVYKGVVNNTTVAVKKLAAM
                    VDITTEELKQOFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVVMPNGSLLDRLSC
                    LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                    ASEKPAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEITIGLPAVDEHRE
                    POLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                    QQLLQEMTAS"
BASE COUNT   463 a   244 c   283 g   393 t
ORIGIN

US-10-001-254-6 (1-98) x AX196260 (1-1383)

Alignment Scores:
Pred. No.:
Score: 513.00
Percent Similarity: 98.98%
Best Local Similarity: 98.98%
Query Match: 99.23%
DB: 2.56e-60

Length: 1383
Matches: 97
Conservative: 0
Mismatches: 1
Indels: 0
Gaps: 0

Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro   20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT   84

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr   40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 144

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 145 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 204

Qy   61 GluLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu   80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 205 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGGATCTTGTGGATCTTG 264

Qy   81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro   98
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 265 ATCCAAAATGAATTTTTTGCTCCTGCGGAGTCTTTTGCTCCCAGATGCTGTTCCC   318

RESULT 9

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 18, 2004, 01:11:43 ; Search time 576.135 Seconds
                                        (without alignments)
                                        2488.940 Million cell updates/sec

Title: US-10-001-254-26
Perfect score: 323
Sequence: 1 MNKPITPSTYVRCLNVGLIR.........KKPSGDDRYNQFHIRCCSQN 59

Scoring table: BLOSUM62
                  Xgapop 10.0 , Xgapext   0.5
                  Ygapop 10.0 , Ygapext   0.5
                  Fgapop   6.0 , Fgapext   7.0
                  Delop   6.0 , Delext   7.0

Searched: 22781392 seqs, 1215223B056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n_model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152424_19723/app_query.fasta_1.1109
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US10001254 @CGN_1_1_4382 @runat_16012004_152424_19723 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
                  1: em_estba:*
                  2: em_esthum:*
                  3: em_estin:*
                  4: em_estmu:*
                  5: em_estov:*
                  6: em_estpl:*
                  7: em_estro:*
                  8: em_htc:*
                  9: gb_est1:*
                 10: gb_est2:*
                 11: gb_htc:*
                 12: gb_est3:*
                 13: gb_est4:*
                 14: gb_est5:*
                 15: em_estun:*
                 16: em_estom:*
                 17: em_gss_hum:*
                 18: em_gss_inv:*
                 19: em_gss_pln:*
                 20: em_gss_vrt:*
                 21: em_gss_fun:*
                 22: em_gss_mam:*
                 23: em_gss_nus:*
                 24: em_gss_pro:*
                 25: em_gss_rod:*
                 26: em_gss_phg:*
                 27: em_gss_vrl:*
                 28: gb_gss1:*

29: gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	323	100.0	284	9	AA114228	AA114228 zn75g05.r1
2	296	91.6	576	10	BF238344	BF238344 601904613
3	290	89.8	540	9	AL699213	AL699213 DKFZp686K
4	290	89.8	719	10	BF696981	BF696981 602130160
5	290	89.8	811	10	BG616438	BG616438 602642772
6	290	89.8	859	10	BG164491	BG164491 602342026
7	284	87.9	313	12	BM151935	BM151935 TCBAP1E11
8	282	87.3	265	10	AW435511	AW435511 168463 BA
9	274	84.8	402	10	BE482619	BE482619 168463 BA
10	271	83.9	453	10	BB860349	BB860349 BB860349
11	271	83.9	503	14	CA538859	CA538859 C0272B01-
12	271	83.9	507	10	BB866698	BB866698 BB866698
13	271	83.9	524	9	AW106160	AW106160 um23h11.y
14	271	83.9	575	4	BX522921	BX522921 H404C09-
15	271	83.9	598	13	BQ552228	BQ552228 BQ552228
16	271	83.9	610	10	BB660378	BB660378 BB660378
17	271	83.9	637	10	BB613447	BB613447 BB613447
18	271	83.9	638	14	BY721552	BY721552 BY721552
19	271	83.9	663	14	BY726858	BY726858 BY726858
20	271	83.9	676	10	BB613167	BB613167 BB613167
21	271	83.9	1161	11	AK020397	AK020397 Mus muscu
22	271	83.9	2481	11	AK028837	AK028837 Mus muscu
23	271	83.9	2810	11	AK029028	AK029028 Mus muscu
24	269	83.3	858	10	BF687921	BF687921 602066996
25	197	61.0	629	9	AL647125	AL647125 AL647125-
26	192	59.4	637	10	BE132064	BE132064 db1h09.y
27	192	59.4	664	12	BJ035962	BJ035962 BJ035962
28	190	58.8	852	13	BU209111	BU209111 603950834
29	186	57.6	670	9	AJ453616	AJ453616 603950834
30	186	57.6	670	9	AJ447581	AJ447581 AJ447581
31	159	49.2	541	10	BG691069	BG691069 34008A BA
32	135	41.8	637	14	CA365604	CA365604 640775 NC
33	125	38.7	555	9	AW423082	AW423082 f168b07.y
34	125	38.7	743	14	CA474136	CA474136 AGENCOURT
35	111	34.4	318	14	CB940191	CB940191 IpCGJx14_
36	100	31.0	380	12	BI883790	BI883790 fn07c09.y
37	74	22.9	477	10	AW958112	AW958112 EST370182
38	73.5	22.8	584	10	BG078473	BG078473 H3028C02-
39	73.5	22.8	588	12	BM022353	BM022353 id59c09.y
40	73.5	22.8	884	12	BI854371	BI854371 603381326
41	71	22.0	508	12	BJ505938	BJ505938 BJ505938
42	71	22.0	605	12	BJ499048	BJ499048 BJ499048
43	69	21.4	631	12	BM138979	BM138979 HC_ad_15D
44	68.5	21.2	778	14	CB236976	CB236976 AGENCOURT
45	68	21.1	803	28	BH443382	BH443382 B0GGL71TR

ALIGNMENTS

RESULT 1
LOCUS   AA114228   284 bp   mRNA   linear   EST 13-NOV-1996
DEFINITION zn75g05.r1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
            cDNA clone IMAGE:564056 5', mRNA sequence.
ACCESSION   AA114228
VERSION   AA114228.1 GI:1668121
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 284)

(handwritten annotations: "DNA encoding seq 26", "AA114228", "seq 26")

Left column

AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE   Generation and analysis of 280,000 human-expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   97044478
PUBMED   8889549

COMMENT   Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham.

FEATURES   Location/Qualifiers
source   1..284
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:4596579"
/db_xref="taxon:9606"
/clone="IMAGE:564056"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/clone_lib="Stratagene NT2 neuronal precursor 937230"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT; Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/c1.D1), Average insert size: 1.0 kb;
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"

BASE COUNT   93 a   63 c   56 g   71 t   1 others
ORIGIN

US-10-001-254-26 (1-59) x AA114228 (1-284)

Alignment Scores:
Pred. No.:
Score: 2.76e-36
                         323.00   Length: 284
Percent Similarity: 100.00%   Matches: 59
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 9   Indels: 0
                                           Gaps: 0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 120

Qy 21 LysLeuSerAspPheIleAspProGlnGlyIleTrpLysLysLeuAlaValAlaIleLeuLys 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AAGCTGTCAGATTTTATTGATCCTCAAGGAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 180

Qy 41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysCysSerGlnAsn 59
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AAACCATCTGGTGATGATAGATACAATCAGTTCACATAAGATGCTGTTCCCAAAAC 237

RESULT 2
BF238344
LOCUS   BF238344   576 bp   mRNA   linear   EST 14-NOV-2000
DEFINITION   601904613F1 NIH_MGC_54 Homo sapiens cDNA clone
IMAGE:4132682 5', mRNA sequence.
ACCESSION   BF238344
VERSION   BF238344.1   GI:11152264
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Right column

REFERENCE   1 (bases 1 to 576)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1032 row: i column: 03
High quality sequence start: 6
High quality sequence stop: 536.

FEATURES   Location/Qualifiers
source   1..576
/organism="Homo sapiens"
/mol_type="mRNA"
/clone="IMAGE:4132682"
/tissue_type="from chronic myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_54"
/note="Organ: bone marrow; Vector: pDNR-lIB (Clontech),
Site 1: SfiI (ggccgcctcggcc); Site 2: SfiI (ggccattatggcc)
; Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACCATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT   191 a   108 c   132 g   145 t
ORIGIN

US-10-001-254-26 (1-59) x BF238344 (1-576)

Alignment Scores:
Pred. No.:
Score: 5.43e-32
                         296.00   Length: 576
Percent Similarity: 100.00%   Matches: 54
Best Local Similarity: 98.18%   Conservative: 1
Query Match: 91.64%   Mismatches: 0
DB: 10   Indels: 0
                                           Gaps: 0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 68 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 127

Qy 21 LysLeuSerAspPheIleAspProGlnGlyIleTrpLysLysLeuAlaValAlaIleLeuLys 40
         |||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db 128 AAGCTGTCAGATTTTATGGATCCTCAAGGAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 187

Qy 41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCys 55
         |||||||||||||||||||||||||||||||||||||||||||
Db 188 AAACCATCTGGTGATGATAGATACAATCAGTTCACATAAGAGC 232

RESULT 3
AL699213
LOCUS   AL699213   540 bp   mRNA   linear   EST 21-MAR-2002
DEFINITION   DKFZp686K18112_r1 686 (synonym: h1cc3) Homo sapiens cDNA clone
DKFZp686K18112 5', mRNA sequence.
ACCESSION   AL699213
VERSION   AL699213.1   GI:19619753
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: January 17, 2004, 23:50:52 ; Search time 51 Seconds
                                           (without alignments)
                                           2544.444 Million cell updates/sec

Title: US-10-001-254-5
Perfect score: 294
Sequence: 1 acatatgtcgcgctgcctcaa.............tgctcccagatgctgttccc 294
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
               1: /cgn2_6/ptcodata/1/ina/5A_COMB.seq:*
               2: /cgn2_6/ptcodata/1/ina/5B_COMB.seq:*
               3: /cgn2_6/ptcodata/1/ina/6A_COMB.seq:*
               4: /cgn2_6/ptcodata/1/ina/6B_COMB.seq:*
               5: /cgn2_6/ptcodata/1/ina/PCTUS_COMB.seq:*
               6: /cgn2_6/ptcodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	100.0	294	833	4	US-09-166-350-10	Sequence 10, Appl
2	40.4	13.7	1879	4	US-09-601-091-1	Sequence 1, Appli
3	40.4	13.7	2024	4	US-09-601-091-3	Sequence 3, Appli
4	40.4	13.7	2024	4	US-09-398-395A-51	Sequence 51, Appl
5	40.4	13.7	2024	4	US-09-887-886A-51	Sequence 51, Appl
6	40.4	13.7	2024	4	US-09-895-752-51	Sequence 51, Appl
7	40.4	13.7	2024	4	US-09-903-012B-51	Sequence 51, Appl
8	33.8	11.5	580073	4	US-08-545-528D-1	Sequence 1, Appl
9	33.6	11.4	2393	4	US-09-734-674-3	Sequence 3, Appl
10	32.8	11.2	202001	4	US-09-961-527-90	Sequence 90, Appl
11	32.6	11.1	1728	4	US-09-096-731A-1	Sequence 1, Appli
12	32.4	11.0	98844	4	US-09-791-211-10	Sequence 10, Appl
13	32.2	11.0	4402	4	US-09-484-970B-135	Sequence 135, App
14	32	10.9	2274	4	US-09-107-532A-3535	Sequence 3535, Ap
15	31.8	10.8	1557	4	US-09-134-001C-1614	Sequence 1614, Ap
16	31.8	10.8	2507	4	US-09-620-312D-120	Sequence 120, App
17	31.8	10.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli
18	31.4	10.7	12720	4	US-08-403-866-11	Sequence 11, Appli
19	31	10.5	2126	2	US-08-545-745-1	Sequence 1, Appli
20	30.8	10.5	1550	4	US-09-427-501-1	Sequence 1, Appl
21	30.8	10.5	15016	4	US-09-601-198-60	Sequence 60, Appl
22	30.8	10.5	319608	4	US-09-539-333D-1	Sequence 1, Appl
23	30.8	10.5	319608	4	US-09-679-409-1	Sequence 1, Appl
24	30.8	10.5	1230025	4	US-09-198-452A-1	Sequence 1, Appli
25	30.6	10.4	1479	4	US-09-328-352-897	Sequence 897, App
26	30.4	10.3	1362	3	US-08-929-329-4	Sequence 4, Appli
27	30.2	10.3	2254	2	US-08-635-066-1	Sequence 1, Appli
28	30	10.2	195	1	US-08-324-243-20	Sequence 20, Appl
29	30	10.2	195	1	US-08-532-390-20	Sequence 20, Appl
30	30	10.2	195	3	US-08-717-294-20	Sequence 20, Appl
31	30	10.2	195	5	PCT-US95-11511-20	Sequence 20, Appl
32	30	10.2	1632	1	US-08-324-243-34	Sequence 34, Appl
33	30	10.2	1632	1	US-08-532-390-34	Sequence 34, Appl
34	30	10.2	1632	3	US-08-717-294-34	Sequence 34, Appl
35	30	10.2	1632	5	PCT-US95-11511-34	Sequence 34, Appl
36	30	10.2	1632	5	PCT-US95-11511-34	Sequence 34, Appl
37	30	10.2	2481	1	US-08-324-243-35	Sequence 35, Appl
38	30	10.2	2481	1	US-08-532-390-35	Sequence 35, Appl
39	30	10.2	2481	3	US-08-717-294-35	Sequence 35, Appl
40	30	10.2	2571	5	PCT-US95-11511-35	Sequence 35, Appl
41	30	10.2	11770	4	US-09-207-914-24	Sequence 24, App
42	30	10.2	4648	4	US-09-961-527-172	Sequence 172, Appl
43	29.8	10.1	4655	4	US-09-207-914-20	Sequence 20, Appl
44	29.6	10.1	1356	1	US-08-592-936B-22	Sequence 22, Appl
45	29.6	10.1	1356	1	US-08-788-928A-2	Sequence 2, Appli

ALIGNMENTS

DNA of Seq ID#5

RESULT 1
US-09-166-350-10

; Sequence 10, Application US/09166350A
; Patent No. 6440663
; GENERAL INFORMATION:
US-09-166-350-10
; APPLICANT: Scanlan, Matthew
; APPLICANT: Chen, Yao
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alex
; TITLE OF INVENTION: Renal Cancer Associated Antigens and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: L0461/7051
; CURRENT APPLICATION NUMBER: US/09/166,350A
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: US 09/166,350
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 833
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-166-350-10

Query Match   100.0%; Score 294; DB 4; Length 833;
Best Local Similarity   100.0%; Pred. No. 3.4e-82;
Matches 294; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ACATATGTCGCGCTGCCTCAATCGTTGGACTAATTGAGGAAGCTGTTCAGAGATTTTATTGATCCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACATATGTCGCGCTGCCTCAATCGTTGGACTAATTGAGGAAGCTGTTCAGAGATTTTATTGATCCT 60

Qy   61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATGATAGATAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATGATAGATAC 133

Qy   121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 193

Qy   181 GAATTACTGTTTGACTGGGGGCACCACAAATTGCACAGTTGGTGGTGATCTTGGGATCTTTTTG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194 GAATTACTGTTTGACTGGGGGCACCACAAATTGCACAGTTGGTGGTGATCTTGGGATCTTTTTG 253

Qy   241 ATCCAAAAATGAAATTTTTTTGCTCCTGCGAGTCTTTTGCTCCCGAGTGCTGTTCCC 294
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314 ATCCAAAAATGAAATTTTTTTGCTCCTGCGAGTCTTTTGCTCCCGAGTGCTGTTCCC 367

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 01:16:08 ; Search time 25.5721 Seconds
(without alignments)
1691.512 Million cell updates/sec

Title: US-10-001-254-6
Perfect score: 517
Sequence: 1 TYVRCLNVGLIRKLSDFIDP..........LLIQNEFFAPASLLPDAVP 98

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152425_19740/app_query.fasta_1.1109
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10001254 @CCN_1_1.95 @runat_16012004_152425_19740 -CPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	517	100.0	833	4	US-09-166-350-10	Sequence 10, Appl
2	96.5	18.7	2288	3	US-09-135-232-1	Sequence 1, Appli
3	96.5	18.7	2288	4	US-09-863-549-1	Sequence 1, Appli
4	79	15.3	1806	2	US-08-980-060-1	Sequence 1, Appli
5	79	15.3	1806	2	US-09-307-185-1	Sequence 1, Appli
6	79	15.3	3459	2	US-08-980-060-3	Sequence 3, Appli
7	79	15.3	3459	3	US-09-307-185-3	Sequence 3, Appli
8	78.5	15.2	479	3	US-08-980-060-14	Sequence 14, Appl
9	78.5	15.2	479	3	US-09-307-185-14	Sequence 14, Appl
10	72	13.9	536165	4	US-09-214-808-1	Sequence 1, Appl
11	69.5	13.4	1416	4	US-09-234-393-26	Sequence 26, Appl
12	69.5	13.4	1416	4	US-09-360-545-19	Sequence 19, Appl
13	69.5	13.4	1416	4	US-09-865-171-26	Sequence 26, Appl
14	69.5	13.4	1785	3	US-09-234-393-49	Sequence 49, Appl
15	69.5	13.4	1785	3	US-09-234-393-51	Sequence 51, Appl
16	69.5	13.4	1785	3	US-09-234-393-53	Sequence 53, Appl
17	69.5	13.4	1785	4	US-09-865-171-49	Sequence 49, Appl
18	69.5	13.4	1785	4	US-09-865-171-51	Sequence 51, Appl
19	69.5	13.4	1785	4	US-09-865-171-53	Sequence 53, Appl
20	69.5	13.4	1785	4	US-09-398-395A-49	Sequence 49, Appl
21	69.5	13.4	1785	4	US-09-887-586A-49	Sequence 49, Appl
22	69.5	13.4	1785	4	US-09-895-752-49	Sequence 49, Appl
23	69.5	13.4	1785	4	US-09-903-012B-49	Sequence 23, Appl
24	69.5	13.4	1977	3	US-09-865-171-23	Sequence 23, Appl
25	66.5	12.9	1977	3	US-09-865-171-23	Sequence 23, Appl
26	66.5	12.9	265	4	US-09-313-294A-3660	Sequence 3660, Ap
27	66	12.8	987	4	US-09-252-991A-11251	Sequence 11251, A
28	66	12.8	1140	4	US-09-252-991A-11181	Sequence 11181, A
29	66	12.8	3141	4	US-09-252-991A-10937	Sequence 10937, A
30	65	12.6	1623	4	US-09-513-783A-33	Sequence 33, Appl
31	64.5	12.5	1521	4	US-09-134-001C-470	Sequence 470, Appl
32	63.5	12.3	1473	3	US-08-907-740-6	Sequence 6, Appli
33	63.5	12.3	3059	4	US-09-620-312D-213	Sequence 213, Appl
34	63.5	12.3	3068	4	US-09-620-312D-214	Sequence 214, App
35	63.5	12.3	3071	4	US-09-620-312D-215	Sequence 215, Appl
36	63.5	12.3	3134	4	US-09-620-312D-212	Sequence 212, App
37	63.5	12.3	3134	4	US-08-743-168B-35	Sequence 35, Appl
38	63	12.2	2472	4	US-08-743-168B-37	Sequence 37, Appl
39	63	12.2	2472	2	US-08-751-189-1	Sequence 1, Appl
40	63	12.2	7881	3	US-08-751-189-1	Sequence 1, Appl
41	63	12.2	7881	3	US-09-184-445-1	Sequence 1, Appli
42	62.5	12.1	1637	4	US-09-724-623-62	Sequence 62, Appli
43	62.5	12.1	4935	3	US-08-631-097-3	Sequence 3, Appl
44	62.5	12.1	5886	3	US-08-810-712-9	Sequence 9, Appl
45	62.5	12.1	6027	2	US-08-968-542C-1	Sequence 1, Appl

ALIGNMENTS

DNA encoding Seq ID #6

RESULT 1
US-09-166-350-10
Sequence 10, Application US/09166350A
Patent No. 6440663 *(circled)*

GENERAL INFORMATION:
APPLICANT: Scanlan, Matthew
APPLICANT: Chen, Yao
APPLICANT: Stockert, Elisabeth
APPLICANT: Old, Lloyd
APPLICANT: Jager, Elke
APPLICANT: Knuth, Alex
TITLE OF INVENTION: Renal Cancer Associated Antigens and
TITLE OF INVENTION: Uses Therefor
FILE REFERENCE: L0461/7051
CURRENT APPLICATION NUMBER: US/09/166,350A
CURRENT FILING DATE: 1998-10-05
EARLIER APPLICATION NUMBER: US 09/166,350
EARLIER FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 10
LENGTH: 833
TYPE: DNA
ORGANISM: Homo sapiens
US-09-166-350-10

Alignment Scores:
Pred. No.: 1.17e-67
Score: 517.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: US-10-001-254-6 (1-98) x US-09-166-350-10 (1-833)

Length: 833
Matches: 98
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
Db   74 ACAATATGTGCGTCGCCTCAATGTTGGACTAATTAGGAGACGTGCAGAGATTTATTGATCT

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
Db 134 CAAGAAGGATGGAGAGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAC 193

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
Db 194 AATCAGTTTCACATAAGGAGGTTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 253

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
Db 254 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTTGGATCTTTTG 313

Qy   81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
Db 314 ATCCAAAATGAATTTTTCCTCCTGCGAGTCTTTTGCTCCCCAGAATGCTGTTCCC 367

RESULT 2
US-09-135-232-1
; Sequence 1, Application US/09135232
; Patent No. 6262228
; GENERAL INFORMATION:
; APPLICANT: Cao, Zhaodan
; TITLE OF INVENTION: IRAK3 Polypeptides, Polynucleotides and Methods
; FILE REFERENCE: T98-019
; CURRENT APPLICATION NUMBER: US/09/135,232
; CURRENT FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 1
; LENGTH: 2288
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(1851)
US-09-135-232-1

Alignment Scores:
Pred. No.:
Score: 0.000283
Percent Similarity: 96.50
Best Local Similarity: 47.95%
Query Match: 31.51%
DB: 18.67%

Length: 2288
Matches: 23
Conservative: 12
Mismatches: 31
Indels: 7
Gaps: 1

US-10-001-254-6 (1-98) x US-09-135-232-1 (1-2288)

Qy   8 ValGlyLeuIleArgLysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeu 27
         :::||| ||: ||| | | || ::
Db 136 CTCGGAGAGCTCTCCGCTGTTCTCGACAGCTGCCACGCGCGCGTGGGCGCCGGCCCTG 195

Qy   28 AlaValAlaIleLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArg 47
         :::
Db 196 GCAGAGAGACTTCAAGC-------------AGCTGGCTGGGAGTGTTCGTCAT 234

Qy   48 PheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGly 67
         ||| :::
Db 235 ATTGAAAAGTATGTAGACCAAGGTAAAAGTGGAACAAGAGAATTACTTTGGTCCTGGGCA 294

Qy   68 ThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
         ||| |||:::|||
Db 295 CAGAAAAACAAGACCATCGGTGACCTTTTACAGGTCCTC 333

RESULT 3
US-09-863-549-1
; Sequence 1, Application US/09863549
; Patent No. 6576444
; GENERAL INFORMATION:
; APPLICANT: Cao, Zhaodan
; TITLE OF INVENTION: IRAK3 Polypeptides, Polynucleotides and Methods
; FILE REFERENCE: T98-019
; CURRENT APPLICATION NUMBER: US/09/863,549
; CURRENT FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: 09/135,232
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 1
; LENGTH: 2288
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(1851)
US-09-863-549-1

Alignment Scores:
Pred. No.:
Score: 0.000283
Percent Similarity: 96.50
Best Local Similarity: 47.95%
Query Match: 31.51%
DB: 18.67%

Length: 2288
Matches: 23
Conservative: 12
Mismatches: 31
Indels: 7
Gaps: 1

US-10-001-254-6 (1-98) x US-09-863-549-1 (1-2288)

Qy   8 ValGlyLeuIleArgLysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeu 27
         :::||| ||: ||| | | || ::
Db 136 CTCGGAGAGCTCTCCGCTGTTCTCGACAGCTGCCACGCGCGCGTGGGCGCCGGCCCTG 195

Qy   28 AlaValAlaIleLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArg 47
         :::
Db 196 GCAGAGAGACTTCAAGC-------------AGCTGGCTGGGAGTGTTCGTCAT 234

Qy   48 PheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGly 67
         ||| :::
Db 235 ATTGAAAAGTATGTAGACCAAGGTAAAAGTGGAACAAGAGAATTACTTTGGTCCTGGGCA 294

Qy   68 ThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
         ||| |||:::|||
Db 295 CAGAAAAACAAGACCATCGGTGACCTTTTACAGGTCCTC 333

RESULT 4
US-08-980-060-1
; Sequence 1, Application US/08980060
; Patent No. 5965421
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: FENG, PING
; APPLICANT: MUZIO, MARTA
; APPLICANT: DIXIT, VISHVA M.
; TITLE OF INVENTION: HUMAN IRAK-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/980,060
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 01:16:08 ; Search time 25.5721 Seconds
               1691.512 Million cell updates/sec

Title:
Perfect score: 517
Sequence: 1 TVVRCLNVGLIRKLSDFIDP...........LLIQNEFPAPASLLLPDAVP 98
Run on: US-10-001-254-6

Scoring table:
               BLOSUM62
               Xgapop 10.0 ,   Xgapext 0.5
               Ygapop 10.0 ,   Ygapext 0.5
               Fgapop 6.0 ,   Fgapext 7.0
               Delop 6.0 ,   Delext 7.0

Searched: 56978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152425_19740/app_query.fasta_1.1109
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10001254 @CGN_1_1_95 @runat_16012004_152425_19740 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -XGAPOP=10 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Issued_Patents_NA:*
1:      /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2:      /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3:      /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4:      /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5:      /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6:      /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	517	100.0	833	4	US-09-166-350-10	Sequence 10, Appl
2	96.5	18.7	2288	3	US-09-135-232-1	Sequence 1, Appli
3	96.5	18.7	2288	3	US-09-863-549-1	Sequence 1, Appli
4	79	15.3	1806	2	US-08-980-060-1	Sequence 1, Appli
5	79	15.3	1806	3	US-09-307-185-1	Sequence 1, Appli
6	79	15.3	3459	3	US-09-980-060-3	Sequence 3, Appli
7	79	15.3	3459	3	US-09-307-185-3	Sequence 3, Appli
8	78.5	15.2	479	2	US-08-980-060-14	Sequence 14, Appl
9	78.5	15.2	479	2	US-09-307-185-14	Sequence 14, Appl
10	72	13.9	536165	4	US-09-214-808-1	Sequence 1, Appli
11	69.5	13.4	1416	3	US-09-234-393-26	Sequence 26, Appl
12	69.5	13.4	1416	4	US-09-360-545-19	Sequence 19, Appl
13	69.5	13.4	1416	4	US-09-865-171-26	Sequence 26, Appl
14	69.5	13.4	1785	3	US-09-234-393-49	Sequence 49, Appl
15	69.5	13.4	1785	3	US-09-234-393-51	Sequence 51, Appl
16	69.5	13.4	1785	3	US-09-234-393-53	Sequence 53, Appl
17	69.5	13.4	1785	4	US-09-865-171-49	Sequence 49, Appl
18	69.5	13.4	1785	4	US-09-865-171-51	Sequence 51, Appl
19	69.5	13.4	1785	4	US-09-865-171-53	Sequence 53, Appl
20	69.5	13.4	1785	4	US-09-398-395A-49	Sequence 49, Appl
21	69.5	13.4	1785	4	US-09-887-586A-49	Sequence 49, Appl
22	69.5	13.4	1785	4	US-09-895-752-49	Sequence 49, Appl
23	69.5	13.4	1785	4	US-09-903-012B-49	Sequence 49, Appl
24	69.5	13.4	1977	3	US-09-234-393-23	Sequence 23, Appl
25	69.5	13.4	1977	4	US-09-865-171-23	Sequence 23, Appl
26	66.5	12.9	265	4	US-09-313-294A-3660	Sequence 3660, Ap
27	66	12.8	987	4	US-09-252-991A-11251	Sequence 11251, A
28	66	12.8	1140	4	US-09-252-991A-11181	Sequence 11181, A
29	66	12.8	3141	4	US-09-252-991A-10937	Sequence 10937, A
30	65	12.6	1623	4	US-09-513-783A-33	Sequence 33, Appl
31	64.5	12.5	1521	4	US-09-134-001C-470	Sequence 470, App
32	63.5	12.3	1473	3	US-08-907-740-6	Sequence 6, Appl
33	63.5	12.3	3059	4	US-09-620-312D-213	Sequence 213, App
34	63.5	12.3	3068	4	US-09-620-312D-214	Sequence 214, App
35	63.5	12.3	3071	4	US-09-620-312D-215	Sequence 215, App
36	63.5	12.3	3134	4	US-09-620-312D-212	Sequence 212, App
37	63	12.2	2472	3	US-08-743-168B-35	Sequence 35, Appl
38	63	12.2	2472	3	US-08-743-168B-37	Sequence 37, Appl
39	63	12.2	7881	2	US-08-751-189-1	Sequence 1, Appli
40	63	12.2	7881	3	US-09-060-036-1	Sequence 1, Appli
41	63	12.2	7881	4	US-09-184-445-1	Sequence 1, Appli
42	62.5	12.1	1637	4	US-09-724-623-62	Sequence 62, Appl
43	62.5	12.1	4935	3	US-09-620-631-3	Sequence 3, Appli
44	62.5	12.1	5886	3	US-08-810-712-9	Sequence 9, Appli
45	62.5	12.1	6027	2	US-08-968-542C-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-166-350-10
Sequence 10, Application US/09166350A
Patent No. 6440663
GENERAL INFORMATION:

APPLICANT: Scanlan, Matthew
APPLICANT: Chen, Yao
APPLICANT: Stockert, Elisabeth
APPLICANT: Old, Lloyd
APPLICANT: Jager, Elke
APPLICANT: Knuth, Alex
TITLE OF INVENTION: Renal Cancer Associated Antigens and
TITLE OF INVENTION: Uses Therefor
FILE REFERENCE: L0461/7051
CURRENT APPLICATION NUMBER: US/09/166,350.
CURRENT FILING DATE: 1998-10-05.
EARLIER APPLICATION NUMBER: US 09/166,350
EARLIER FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 10
LENGTH: 833
TYPE: DNA
ORGANISM: Homo sapiens
US-09-166-350-10

DNA encoding Seq ID #6

Alignment Scores:
Pred. No.:   1.17e-67
Score: 517.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB:

Length: 833
Matches: 98
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-001-254-6 (1-98) x US-09-166-350-10 (1-833)

Sequence 26, Appl1
Sequence 49, Appl
Sequence 51, Appl
Sequence 53, Appl
Sequence 49, Appl
Sequence 51, Appl
Sequence 53, Appl
Sequence 49, Appl
Sequence 49, Appl
Sequence 49, Appl
Sequence 49, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 3660, Ap
Sequence 11251, A
Sequence 11181, A
Sequence 10937, A
Sequence 33, Appl
Sequence 470, App
Sequence 6, Appl
Sequence 213, App
Sequence 214, App
Sequence 215, App
Sequence 212, App
Sequence 35, Appl
Sequence 37, Appl
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 62, Appl
Sequence 3, Appli
Sequence 9, Appli
Sequence 1, Appl

Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgGlyLeuSerAspPheIleAspPro 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74 ACATATGTGCGCTGCCTCAATGTGGACTAATTAGGAGCGTCGACGATTTTATTGATCCT 133

Qy   21 GlnGluGlyTrpLysLysProSerGlyAspAspArgTyr 40
        ||||||||||||||||||||||||||||||||||||||||
Db 134 CAAGAAGGATGGAAGAAGAGTTAGCTGTACCATTAAAAAACCATCTGGTGATGATAGATAC 193

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 194 AATCAGTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 253

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 313

Qy   81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 314 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 367

Alignment Scores:
Pred. No.:   Length: 2288
Score: 0.000283   Matches: 23
Percent Similarity: 47.95%   Conservative: 12
Best Local Similarity: 31.51%   Mismatches: 31
Query Match: 18.67%   Indels: 7
DB: Gaps: 1

US-10-001-254-6 (1-98) x US-09-863-549-1 (1-2288)

Qy   8 ValGlyLeuIleArgGlyLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeu 27
        ::::||||||||||||||||:::: |||:||::
Db 136 CTCGGAGAGCTCTCGGCCTGTTCTGGACAGCTGCCGACGGCGGGGCGGGGGCTG 195

Qy   28 AlaValAlaIleLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArg 47
        ||| ::: |||
Db 196 GCAGAGAGACTTTCAAGC----------------------AGCTGGCTGGATGTTCGTCAT 234

Qy   48 PheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGly 67
        ||| ||| :::
Db 235 ATTGAAAAGTATGTAGACCAAGGTAAAAGTGGAACAAGAGAATTACTTTGGTCCTGGGCA 294

Qy   68 ThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
        ||| :::   :::|||
Db 295 CAGAAAAACAAGACCATCGGTGACCTTTTACAGGTCCTC 333

RESULT 3
US-09-863-549-1
; Sequence 1, Application US/09863549
; Patent No. 6576444
; GENERAL INFORMATION:
; APPLICANT: Cao, Zhaodan

RESULT 2
US-09-135-232-1
; Sequence 1, Application US/09135232
; Patent No. 6262228
; GENERAL INFORMATION:
; APPLICANT: Cao, Zhaodan

Alignment Scores:
Pred. No.:   Length: 2288
Score: 0.000283   Matches: 23
Percent Similarity: 47.95%   Conservative: 12
Best Local Similarity: 31.51%   Mismatches: 31
Query Match: 18.67%   Indels: 7
DB: Gaps: 1

US-10-001-254-6 (1-98) x US-09-135-232-1 (1-2288)

Qy   8 ValGlyLeuIleArgGlyLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeu 27
        ::::||||||||||||||||:::: |||:||::
Db 136 CTCGGAGAGCTCTCGGCCTGTTCTGGACAGCTGCCGACGGCGGGGCGGGGGCTG 195

Qy   28 AlaValAlaIleLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArg 47
        ||| ::: |||
Db 196 GCAGAGAGACTTTCAAGC----------------------AGCTGGCTGGATGTTCGTCAT 234

Qy   48 PheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGly 67
        ||| ||| :::
Db 235 ATTGAAAAGTATGTAGACCAAGGTAAAAGTGGAACAAGAGAATTACTTTGGTCCTGGGCA 294

Qy   68 ThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
        ||| :::   :::|||
Db 295 CAGAAAAACAAGACCATCGGTGACCTTTTACAGGTCCTC 333

RESULT 4
US-08-980-060-1
; Sequence 1, Application US/08980060
; Patent No. 5965421
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: FENG, PING
; APPLICANT: RUZIO, MARTA
; APPLICANT: DIXIT, VISHVA M.
; TITLE OF INVENTION: HUMAN IRAK-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER: IBM PC compatible
; MEDIUM TYPE: Floppy disk
; COMPUTER READABLE FORM:
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/980,060
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1806 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 34..1803

US-08-980-060-1

Alignment Scores:
Pred. No.:
Score: 0.0822
Percent Similarity: 79.00
Best Local Similarity: 49.28%
Query Match: 31.88%
DB: 15.28%
                                2

Length: 1806
Matches: 22
Conservative: 12
Mismatches: 19
Indels: 16
Gaps: 2

US-10-001-254-6 (1-98) x US-08-980-060-1 (1-1806)

```
Qy    45  IleArgArgPheGluAlaLeuLeuGlnThr---GlyLysSerProThrSerGluLeuLeu  63
              :::||||:::  :::                    |||||  ::::|:
Db   157  CTGCGGAAGATCAAGTTCATGGAGCGGGTGTGAGCATCACGCGGGAGCTGCTG  216

Qy    64  PheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsn  83
              :::  |||||||              |||||||||||||||:::
Db   217  TGGTGGTGGGGCCATGCGGCAGGCCACCGTCCAGCAACTTGTGGACTTCCTGTGCCGCCTG  276

Qy    84  GluPhePheAlaProAlaSerLeuLeuLeu------------------------  93
              |||  ::::  |||
Db   277  GAGCTCTACCGGGCTGCCCAGATCATCCTGAACTGGAAACCGGCTCCTGAAATCAGGTGT  336

Qy    94  ------------ProAspAlaVal  97
                          |||||:::|||
Db   337  CCCATTCCAGCCTTCCCTGACTCTGTG  363
```

RESULT 5
US-09-307-185-1
Sequence 1, Application US/0930718S
Patent No. 6222019
GENERAL INFORMATION:
APPLICANT: NI, JIAN
APPLICANT: FENG, PING
APPLICANT: MUZIO, MARTA
APPLICANT: DIXIT, VISHVA M.
TITLE OF INVENTION: HUMAN IRAK-2
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/307,185
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/980,060
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1806 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 34..1803

US-09-307-185-1

Alignment Scores:
Pred. No.:
Score: 0.0822
Percent Similarity: 79.00
Best Local Similarity: 49.28%
Query Match: 31.88%
DB: 15.28%
                                3

Length: 1806
Matches: 22
Conservative: 12
Mismatches: 19
Indels: 16
Gaps: 2

US-10-001-254-6 (1-98) x US-09-307-185-1 (1-1806)

```
Qy    45  IleArgArgPheGluAlaLeuLeuGlnThr---GlyLysSerProThrSerGluLeuLeu  63
              :::||||:::  :::                    |||||  ::::|:
Db   157  CTGCGGAAGATCAAGTTCATGGAGCGGGTGTGAGCATCACGCGGGAGCTGCTG  216

Qy    64  PheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsn  83
              :::  |||||||              |||||||||||||||:::
Db   217  TGGTGGTGGGGCCATGCGGCAGGCCACCGTCCAGCAACTTGTGGACTTCCTGTGCCGCCTG  276

Qy    84  GluPhePheAlaProAlaSerLeuLeuLeu------------------------  93
              |||  ::::  |||
Db   277  GAGCTCTACCGGGCTGCCCAGATCATCCTGAACTGGAAACCGGCTCCTGAAATCAGGTGT  336

Qy    94  ------------ProAspAlaVal  97
                          |||||:::|||
Db   337  CCCATTCCAGCCTTCCCTGACTCTGTG  363
```

RESULT 6
US-08-980-060-3
Sequence 3, Application US/08980060
Patent No. 5965421
GENERAL INFORMATION:
APPLICANT: NI, JIAN
APPLICANT: FENG, PING
APPLICANT: MUZIO, MARTA
APPLICANT: DIXIT, VISHVA M.
TITLE OF INVENTION: HUMAN IRAK-2
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/980,060
FILING DATE: Herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.

; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3459 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..1908
US-08-980-060-3

Alignment Scores:
Pred. No.:
Score: 0.211
Percent Similarity: 79.00
Best Local Similarity: 49.28%
Query Match: 31.88%
DB: 15.28%

Length: 3459
Matches: 22
Conservative: 12
Mismatches: 19
Indels: 16
Gaps: 2

US-10-001-254-6 (1-98) x US-08-980-060-3 (1-3459)

Qy   45 IleArgArgPheGluAlaLeuLeuGlnThr---GlyLysSerProThrSerGluLeuLeu 63
             ::::|||||| ::::::::::::: ||| ||||||||
Db   157 CTGCGGAAGATCAAGTCCATGGAGCCGGGTGCAGGGTGTGAGCATCACGCGGGAGCTGCTG 216

Qy   64 PheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsn 83
             :::|||| :::|||||| :::
Db   217 TGGTGGTGGGGCATGCGGGACAGGCCACCGGTCCAGCAACTTGTGGACCTCCTTGTGCCGCCTG 276

Qy   84 GluPhePheAlaProAlaSerLeuLeuLeu-------------------------------- 93
             ::: |||
Db   277 GAGCTCTACCGGGCGTGCCCAGATCATCCTGAAACTGGAAACCGGCTTCCTGAAATCAGGTGT 336

Qy   94 -------------------ProAspAlaVal 97
                                ||||||:::|||
Db   337 CCCATTCCAGCCTTCCCTGACTCTGTG 363

RESULT 7
US-09-307-185-3
; Sequence 3, Application US/09307185
; Patent No. 6222019
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: FENG, PING
; APPLICANT: MUZIO, MARTA
; APPLICANT: DIXIT, VISHVA M.
; TITLE OF INVENTION: HUMAN IRAK-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/307,185
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/980,060
; FILING DATE:

ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3459 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..1908
US-09-307-185-3

Alignment Scores:
Pred. No.:
Score: 0.211
Percent Similarity: 79.00
Best Local Similarity: 49.28%
Query Match: 31.88%
DB: 15.28%

Length: 3459
Matches: 22
Conservative: 12
Mismatches: 19
Indels: 16
Gaps: 2

US-10-001-254-6 (1-98) x US-09-307-185-3 (1-3459)

Qy   45 IleArgArgPheGluAlaLeuLeuGlnThr---GlyLysSerProThrSerGluLeuLeu 63
             ::::|||||| ::::::::::::: ||| ||||||||
Db   157 CTGCGGAAGATCAAGTCCATGGAGCCGGGTGCAGGGTGTGAGCATCACGCGGGAGCTGCTG 216

Qy   64 PheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsn 83
             :::|||| :::|||||| :::
Db   217 TGGTGGTGGGGCATGCGGGACAGGCCACCGGTCCAGCAACTTGTGGACCTCCTTGTGCCGCCTG 276

Qy   84 GluPhePheAlaProAlaSerLeuLeuLeu-------------------------------- 93
             ::: |||
Db   277 GAGCTCTACCGGGCGTGCCCAGATCATCCTGAAACTGGAAACCGGCTTCCTGAAATCAGGTGT 336

Qy   94 -------------------ProAspAlaVal 97
                                ||||||:::|||
Db   337 CCCATTCCAGCCTTCCCTGACTCTGTG 363

RESULT 8
US-08-980-060-14/c
; Sequence 14, Application US/08980060
; Patent No. 5965421
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: FENG, PING
; APPLICANT: MUZIO, MARTA
; APPLICANT: DIXIT, VISHVA M.
; TITLE OF INVENTION: HUMAN IRAK-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/980,060
; FILING DATE: Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

US-08-980-060-14

Alignment Scores:
Pred. No.:
Score: 0.0143
Percent Similarity: 78.50
Best Local Similarity: 60.00%
Query Match: 38.00%
                          15.18%
                          2

Length: 479
Matches: 19
Conservative: 11
Mismatches: 19
Indels: 1
Gaps: 1

US-10-001-254-6 (1-98) × US-08-980-060-14 (1-479)

```
Qy   45  IleArgArgPheGluAlaLeuLeuGlnThr---GlyLysSerProThrSerGluLeuLeu  63
             :::||:||:||  :::  ||||||||||
Db  244  CTGCGGAAGATCAAGTCCATGGAGCGCGGGTGCAGGGGTGTGACCATCACGCGGGAGCTGCTG  185

Qy   64  PheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsn  83
             :::
Db  184  TGGTGGTGGGGGCATGGGGAGGCACCACGGTCAGCAACTTGTGGACTTCCTCCTGTGCCCGTG  125

Qy   84  GluPheAlaProAlaSerLeuLeuLeu  93
             |||  :::::||:|||
Db  124  GAGCTCTACCGGGCGCTGCCCAGATCATCCTG  95
```

RESULT 9
US-09-307-185-14/c
; Sequence 14, Application US/0930718 5
; Patent No. 6222019
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: FENG, PING
; APPLICANT: MUZIO, MARTA
; APPLICANT: DIXIT, VISHVA M.
; TITLE OF INVENTION: HUMAN IRAK-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/307,185
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/980,060
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

US-09-307-185-14

Alignment Scores:
Pred. No.:
Score: 0.0143
Percent Similarity: 78.50
Best Local Similarity: 60.00%
Query Match: 38.00%
                          15.18%
                          3

Length: 479
Matches: 19
Conservative: 11
Mismatches: 19
Indels: 1
Gaps: 1

US-10-001-254-6 (1-98) × US-09-307-185-14 (1-479)

```
Qy   45  IleArgArgPheGluAlaLeuLeuGlnThr---GlyLysSerProThrSerGluLeuLeu  63
             :::||:||:||  :::  ||||||||||
Db  244  CTGCGGAAGATCAAGTCCATGGAGCGCGGGTGCAGGGGTGTGACCATCACGCGGGAGCTGCTG  185

Qy   64  PheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsn  83
             :::
Db  184  TGGTGGTGGGGGCATGGGGAGGCACCACGGTCAGCAACTTGTGGACTTCCTCCTGTGCCCGTG  125

Qy   84  GluPheAlaProAlaSerLeuLeuLeu  93
             |||  :::::||:|||
Db  124  GAGCTCTACCGGGCGCTGCCCAGATCATCCTG  95
```

RESULT 10
US-09-214-808-1
; Sequence 1, Application US/09214808A
; Patent No. 6475793
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic plasmid
; FILE REFERENCE: CARP0068
; CURRENT APPLICATION NUMBER: US/09/214,808A
; CURRENT FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: PCT/IB97/00950
; PRIOR FILING DATE: 1997-07-10
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium

US-09-214-808-1

Alignment Scores:
Pred. No.:
Score: 3.46e+03
Percent Similarity: 72.00
Best Local Similarity: 47.06%
Query Match: 32.94%
                          13.93%
                          4

Length: 536165
Matches: 28
Conservative: 12
Mismatches: 39
Indels: 6
Gaps: 5

US-10-001-254-6 (1-98) × US-09-214-808-1 (1-536165)

```
Qy    18  IleAspProGlnGlyGlyTrp---LysLysLeuAlaValAlaIleLysLysGlyProSer---  35
              |||  :::|||  |||  :::|| ||||||||
Db 97239  ATAGACGATCGGGAGGACTGGGACACAGAGAAAAACATGGCTCGTCACGTCTCCGTCTGCC  97298

Qy    36  ---GlyAspAspArg---TyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeuGln  53
              |||||||||   :::
```

Db 97299 ACCGGGCGATGACATCTGACTTGGCTCCCTGCGAGATCCACCAGATAATCAAATCTTGTTTAG 97358

Qy 54 ThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThrVal 73

Db 97359 ATTAGGCGCTTCGGCTCAACTTAGATTGCTTCTTGGATGG---TCTACTAACTGCCCTCAA 97415

Qy 74 --GlyAspLeuValAspLeuLeuIleGlnAsnGluPheAlaProAlaSerLeuLeu 92

Db 97416 TGAAGTGACCTTACAGGCCGGGCGGCTCTTTCCAAATCCGCTTCTTATCCGCTCTTTCAAAATTG 97475

Qy 93 LeuProAspAlaVal 97
        :::
Db 97476 GTCCAATCAAATTGTT 97490

RESULT 11
US-09-234-393-26
; Sequence 26, Application US/09234393A
; Patent No. 6265539
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Crock, John E
; APPLICANT: Bohlman, Jorg
; APPLICANT: Jetter, Reinhard
; APPLICANT: Steele, Christopher L
; TITLE OF INVENTION: SESQUITERPENE SYNTHASES FROM GRAND FIR (ABIES GRANDIS)
; TITLE OF INVENTION: AND METHODS OF USE
; FILE REFERENCE: WSUR11334S
; CURRENT APPLICATION NUMBER: US/09/234,393A
; CURRENT FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: 60/072,204
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Abies grandis
US-09-234-393-26

Alignment Scores:
Pred. No.:
Score: 1.51
Percent Similarity: 69.50
Best Local Similarity: 40.91%
Query Match: 31.06%
DB: 13.44%

Length: 1416
Matches: 41
Conservative: 13
Mismatches: 37
Indels: 41
Gaps: 10

US-10-001-254-6 (1-98) x US-09-234-393-26 (1-1416)

Qy 3 ValArgCysGluLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAsp---ProGln 21
        ::||||| ||| ::|||||||| ||||::
Db 501 GTGAGACGATGGGAGTTTCGTTCGGTAGAGGGCCTCCCAGACTTCATGAAAAATGCATTC 560

Qy 22 GluGlyTrpLysLys--------------LeuAlaValAlaIleLeuLysGlyProSerGly 36
         ||||| ::||| |||||||||
Db 561 GAGTTCTGGTTAAAGACATCTAATGAATTGGTCAAGCTGTT---AAAGCCGCAAGGG 617

Qy 37 AspAsp---------------------ArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeu 51
         ||| ::||| |||||||
Db 618 CAAGATATGGCGGGCCTACATAAGAAAAATGCATGGGAG---CGATACCTTGAAGCTTAT 674

Qy 52 LeuGln-----------------------ThrGlyThrLysSerProThrSerGluLeuLeuPheAsp 65
         ||||| :::
Db 675 CTGCAAGATGCGGAATGGATAGCCACTGGATGTGTGTATTG---AATTTGATTCCGCTTCTTAATG 734

Qy 66 TrpGlyThrThrAsn-------------CysThrValGlyAspLeuValAspLeuLeuIleGln 82
         ||| :::
Db 735 AATGGCACCACCAAACACTGGGATGGTGTGTATTG---AATTGATTCCGCTTCTTCGTTAATG 791

Qy 83 AsnGlu-----------------------PhePheAlaPro--------------- 88
         |||
Db 792 GGTGAACATTTACCAAATCGACATTCTCGGGACGCAAATATTCTTGCCCTCCAGGTTCCACCAT 851

US-10-001-254-6.rni

Qy 89 ---------------------AlaSerLeuLeuProAspAla 96
         ||||| |||::| |||||
Db 852 CTCATTGAATTGGCTTCCAGGCTCGTCGGATGACGCG 887

RESULT 12
US-09-360-545-19
; Sequence 19, Application US/09360545
; Patent No. 6429014
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Bohlmann, Jorg
; APPLICANT: Steele, Christopher L
; APPLICANT: Phillips, Michael A
; TITLE OF INVENTION: MONOTERPENE SYNTHASES FROM GRAND FIR (ABIES GRANDIS)
; FILE REFERENCE: wsur13885
; CURRENT APPLICATION NUMBER: US/09/360,545
; CURRENT FILING DATE: 1999-07-26
; EARLIER APPLICATION NUMBER: 60/052,249
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: PCT/US98/14528
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Abies grandis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(1199)
; OTHER INFORMATION: Clone AG5.9
US-09-360-545-19

Alignment Scores:
Pred. No.:
Score: 1.51
Percent Similarity: 69.50
Best Local Similarity: 40.91%
Query Match: 31.06%
DB: 13.44%

Length: 1416
Matches: 41
Conservative: 13
Mismatches: 37
Indels: 41
Gaps: 10

US-10-001-254-6 (1-98) x US-09-360-545-19 (1-1416)

Qy 3 ValArgCysGluLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAsp---ProGln 21
        ::||||| ||| ::|||||||| ||||::
Db 501 GTGAGACGATGGGAGTTTCGTTCGGTAGAGGGGCCTCCCAGACTTCATGAAAAATGCATTC 560

Qy 22 GluGlyTrpLysLys--------------LeuAlaValAlaIleLeuLysGlyProSerGly 36
         ||||| ::||| |||||||||
Db 561 GAGTTCTGGTTAAAGACATCTAATGAATTGGTCAAGCTGTT---AAAGCCGCAAGGG 617

Qy 37 AspAsp---------------------ArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeu 51
         ||| ::||| |||||||
Db 618 CAAGATATGGCGGGCCTACATAAGAAAAATGCATGGGAG---CGATACCTTGAAGCTTAT 674

Qy 52 LeuGln-----------------------ThrGlyThrLysSerProThrSerGluLeuLeuPheAsp 65
         ||||| :::
Db 675 CTGCAAGATGCGGAATGGATAGCCACTGGATGTGTGTATTG---AATTTGATTCCGCTTCTTAATG 734

Qy 66 TrpGlyThrThrAsn-------------CysThrValGlyAspLeuValAspLeuLeuIleGln 82
         ||| :::
Db 735 AATGGCACCACCAAACACTGGGATGGTGTGTATTG---AATTGATTCCGCTTCTTCGTTAATG 791

Qy 83 AsnGlu-----------------------AlaSerLeuLeuProAspAla 96
         |||
Db 792 GGTGAACATTTACCAAATCGACATTCTCGGGACGCAAATATTCTTGCCCTCCAGGTTCCACCAT 851

Qy 89 --------------AlaSerLeuLeuProAspAla 96
         |||::| |||||
Db 852 CTCATTGAATTGGCTTCCAGGCTCGTCGGATGACGCG 887

RESULT 13
US-09-865-171-26

Sequence 26, Application US/09865171
; Patent No. 6451576
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Crock, John E
; APPLICANT: Bohlmann, Jorg
; APPLICANT: Steele, Christopher L
; TITLE OF INVENTION: SESQUITERPENE SYNTHASES FROM GRAND FIR (ABIES GRANDIS),
; TITLE OF INVENTION: AND METHODS OF USE
; FILE REFERENCE: WSUR117468
; CURRENT APPLICATION NUMBER: US/09/865,171
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 09/234,393
; PRIOR FILING DATE: 1999-01-20
; PRIOR APPLICATION NUMBER: 60/072,204
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Abies grandis

US-09-865-171-26

Alignment Scores:
Pred. No.:   1.51
Score: 69.50
Percent Similarity: 40.91%
Best Local Similarity: 31.06%
Query Match: 13.44%

Length: 1416
Matches: 41
Conservative: 13
Mismatches: 37
Indels: 41
Gaps: 10
DB:

US-10-001-254-6 (1-98) x US-09-865-171-26 (1-1416)

```
Qy    3 ValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAsp---ProGln 21
          ||           |||                |||||||    |||
Db  501 GTGAGACGATGGGATGTTTCGTTGGTTAGAGGGCGCTCCCAGACTTCATTGAAAATTGCATTC 560

Qy   22 GluGlyTrpLysLys----------LeuAlaValAlaIleLeuLysProSerGly 36
          ||| |||                   ||           |||
Db  561 GAGTTCTGGTTAAAGACATCTAATGAATTGCTGAAGCTGTT---AAAGCGCAAGGG 617

Qy   37 AspAsp-----------ArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeu 51
          |||                ||| 
Db  618 CAAGATATGGGCGGGCCTTACATAAGAAAAAATGCATGGGAG---CGATACCTTGAAGCTTAT 674

Qy   52 LeuGln------------ThrGlyLysSerProThrSerGluLeuLeuPheAsp 65
          ||||||             |||
Db  675 CTGCAAGATGCGGGAATGGAATAGCCACTGGACAATGTCCCCACCTTTGATGAGTACTTGAAT 734

Qy   66 TrpGlyThrThrAsn------------CysThrValGlyAspLeuValAspLeuLeuIleGln 82
          ||| |||                    :::       |||:::|||
Db  735 AATGGCACACCAAACACTGGGGATGTGTGTAATG---AATTTGATTCCGCTTCTTGTTAATG 791

Qy   83 AsnGlu----------------PhePheAlaPro-------------- 88
          |||                   |||
Db  792 GGTGAACATTTACCAATCGACATTCTGGAGCAAAATATTCTTGCCCTCCAGGTTCCACCAT 851

Qy   89 ------------AlaSerLeuLeuProAspAlaAla 96
                       ||| :::   |||||
Db  852 CTCATTGAATTGGCTTCCAGGCTCGTTCGATGACGCG 887
```

RESULT 14
US-09-234-393-49
; Sequence 49, Application US/0924393A
; Patent No. 6265639
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Crock, John E
; APPLICANT: Bohlmann, Jorg
; APPLICANT: Jetter, Reinhard
; APPLICANT: Steele, Christopher L
; TITLE OF INVENTION: SESQUITERPENE SYNTHASES FROM GRAND FIR (ABIES GRANDIS)
; TITLE OF INVENTION: AND METHODS OF USE
; FILE REFERENCE: WSUR113345
; CURRENT APPLICATION NUMBER: US/09/234,393A
; CURRENT FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: 60/072,204
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: computer-generated nucleic acid sequence encoding
; OTHER INFORMATION: gamma humulene synthase
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4)..(1782)

US-09-234-393-49

Alignment Scores:
Pred. No.:   2.11
Score: 69.50
Percent Similarity: 40.91%
Best Local Similarity: 31.06%
Query Match: 13.44%

Length: 1785
Matches: 41
Conservative: 13
Mismatches: 37
Indels: 41
Gaps: 10
DB:

US-10-001-254-6 (1-98) x US-09-234-393-49 (1-1785)

```
Qy    3 ValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAsp---ProGln 21
          ||           |||                |||||||    |||
Db 1087 GTGAGACGATGGGATGTTTCGTTGGTTAGAGGGCGCTCCCAGACTTCATTGAAAATTGCATTC 1146

Qy   22 GluGlyTrpLysLys----------LeuAlaValAlaIleLeuLysProSerGly 36
          ||| |||                   ||           |||
Db 1147 GAGTTCTGGTTAAAGACATCTAATGAATTGCTGAAGCTGTT---AAAGCGCAAGGG 1203

Qy   37 AspAsp-----------ArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeu 51
          |||                ||| 
Db 1204 CAAGATATGGGCGGGCCTTACATAAGAAAAAATGCATGGGAG---CGATACCTTGAAGCTTAT 1260

Qy   52 LeuGln------------ThrGlyLysSerProThrSerGluLeuLeuPheAsp 65
          ||||||             |||
Db 1261 CTGCAAGATGCGGGAATGGAATAGCCACTGGACAATGTCCCCACCTTTGATGAGTACTTGAAT 1320

Qy   66 TrpGlyThrThrAsn------------CysThrValGlyAspLeuValAspLeuLeuIleGln 82
          ||| |||                    :::       |||:::|||
Db 1321 AATGGCACACCAAACACTGGGGATGTGTGTAATG---AATTTGATTCCGCTTCTTGTTAATG 1377

Qy   83 AsnGlu----------------PhePheAlaPro-------------- 88
          |||                   |||
Db 1378 GGTGAACATTTACCAATCGACATTCTGGAGCAAAATATTCTTGCCCTCCAGGTTCCACCAT 1437

Qy   89 ------------AlaSerLeuLeuProAspAlaAla 96
                       ||| :::   |||||
Db 1438 CTCATTGAATTGGCTTCCAGGCTCGTTCGATGACGCG 1473
```

RESULT 15
US-09-234-393-51
; Sequence 51, Application US/0924393A
; Patent No. 6265639
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Crock, John E
; APPLICANT: Bohlmann, Jorg
; APPLICANT: Jetter, Reinhard
; APPLICANT: Steele, Christopher L
; TITLE OF INVENTION: SESQUITERPENE SYNTHASES FROM GRAND FIR (ABIES GRANDIS)
; TITLE OF INVENTION: AND METHODS OF USE
; FILE REFERENCE: WSUR113345
; CURRENT APPLICATION NUMBER: US/09/234,393A

```
;   CURRENT FILING DATE: 1999-01-20
;   EARLIER APPLICATION NUMBER: 60/072,204
;   EARLIER FILING DATE: 1998-01-22
;   NUMBER OF SEQ ID NOS: 55
;   SOFTWARE: Patentin Ver. 2.0
;
;   SEQ ID NO 51
;     LENGTH: 1785
;     TYPE: DNA
;     ORGANISM: Artificial Sequence
;   FEATURE:
;     OTHER INFORMATION: Description of Artificial Sequence:
;     OTHER INFORMATION: computer-generated nucleic acid sequence encoding
;     OTHER INFORMATION: gamma humulene synthase
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: (4)..(1782)
;   US-09-234-393-51

Alignment Scores:
Pred. No.:              2.11         Length:         1785
Score:                 69.50         Matches:          41
Percent Similarity:    40.91%        Conservative:     13
Best Local Similarity: 31.06%        Mismatches:       37
Query Match:           13.44%        Indels:           41
DB:                       3          Gaps:             10

US-10-001-254-6 (1-98) x US-09-234-393-51 (1-1785)

Qy      3 ValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAsp---ProGln 21
            ||||            |||||           |||||        ||:||
Db   1087 GTGAGACGATGGGGATGGTTTCGTTCGTTGGTTAGAGGGCCTTCCCAGACTTCATGAAAATTGCATTC 1146

Qy     22 GluGlyTrpLysLys----------------LeuAlaValAlaIleLysLysProSerGly 36
            ||||| |||||                     |||::::      |||          |||
Db   1147 GAGTTCTGGTTAAAGACATCTAATCAATCGAAATTGATTGCTGAAGCTGTT---AAAGCGCAAGG 1203

Qy     37 AspAsp-------------ArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeu 51
            |||                   |||                   |||||
Db   1204 CAAGATATGGGGCGGGCCTACATAAGAAAAAATGCATGGGAG---CGATACCTTGAAGCTTAT 1260

Qy     52 LeuGln--------------ThrGlyLysSerProThrSerGluLeuLeuPheAsp 65
            |||||                ||||||                          :::
Db   1261 CTGCAAGATGCGGAATGGGAATAGCCACTGGACATGTCCCCACCTTTGATGAGTACTTGAAT 1320

Qy     66 TrpGlyThrThrAsn--------CysThrValGlyAspLeuValAspLeuLeuIleGln 82
            ||||| |||      |||              |||          |||:::
Db   1321 AATGGCACACCAAAACACTGGGATGTGTATTG--AATTTGATTCCGCTTCTTGTTAATG 1377

Qy     83 AsnGlu--------------PhePheAlaPro--------------- 88
            |||                   |||
Db   1378 GGTGAACATTTACCAATCGACATTCTGGAGCAAATATTTCTTGCCCTTCCCAGGTTCCACCAT 1437

Qy     89 ---------AlaSerLeuLeuLeuProAspAla 96
                      ||||||           ||:||
Db   1438 CTCATTGAATTGGGCTTCCAGGCTCCGTCGATGACGCG 1473
```

```
OM protein - nucleic search, using frame_plus_p2n model

                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:             January 18, 2004, 01:04:18 ; Search time 1658.38 Seconds
                                                  (without alignments)
                                                  2417.512 Million cell updates/sec

Title:              US-10-001-254-6
Perfect score:      517
Sequence:           1 TVVRCLNVGLIRKLSDFIDP..........LLIQNEFPAPASLLLPDAVP 98

Scoring table:      BLOSUM62
                    Xgapop 10.0 ,  Xgapext  0.5
                    Ygapop 10.0 ,  Ygapext  0.5
                    Fgapop  6.0 ,  Fgapext  7.0
                    Delop   6.0 ,  Delext   7.0

Searched:           2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters:        5777422

Minimum DB seq length:   0
Maximum DB seq length:   200000000

Post-processing:    Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Command line parameters:
 -MODEL=frame+_p2n.model  -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152423_19714/app_query.fasta_1.1109
 -Db=GenEmbl  -QFMT=fastap  -SUFFIX=rge  -MINMATCH=0.1  -LOOPCL=0  -LOOPEXT=0
 -UNITS=bits  -START=1  -END=-1  -MATRIX=blosum62  -TRANS=human40.cdi  -LIST=45
 -DOCALIGN=200  -THR_SCORE=pct  -THR_MAX=100  -THR_MIN=0  -ALIGN=15  -MODE=LOCAL
 -OUTFMT=pto  -NORM=ext  -HEAPSIZE=500  -MINLEN=0  -MAXLEN=200000000000
 -USER=US10001254_@CGN_1_5671 @runat_16012004_152423_19714  -NCPU=6  -ICPU=3
 -NO_MAP  -LARGEQUERY  -NEG_SCORES=0  -WAIT  -DSPBLOCK=100  -LONGLOG
 -DEV_TIMEOUT=120  -WARN_TIMEOUT=30  -THREADS=1  -XGAPOP=10  -XGAPEXT=0.5  -FGAPOP=6
 -FGAPEXT=7  -YGAPOP=10  -YGAPEXT=0.5  -DELOP=6  -DELEXT=7

Database :          GenEmbl:*

                    1:  gb_ba:*
                    2:  gb_htg:*
                    3:  gb_in:*
                    4:  gb_om:*
                    5:  gb_ov:*
                    6:  gb_pat:*
                    7:  gb_ph:*
                    8:  gb_pl:*
                    9:  gb_pr:*
                   10:  gb_ro:*
                   11:  gb_sts:*
                   12:  gb_sy:*
                   13:  gb_un:*
                   14:  gb_vi:*
                   15:  gb_ba:*
                   16:  gb_fun:*
                   17:  gb_hum:*
                   18:  gb_in:*
                   19:  gb_mu:*
                   20:  gb_om:*
                   21:  gb_or:*
                   22:  gb_ov:*
                   23:  gb_pat:*
                   24:  gb_ph:*
                   25:  gb_pl:*
                   26:  gb_ro:*
                   27:  gb_sts:*
                   28:  gb_un:*
                   29:  em_vi:*
                   30:  em_htg_hum:*
                   31:  em_htg_inv:*
                   32:  em_htg_other:*
                   33:  em_htg_mus:*
                   34:  em_htg_pln:*
                   35:  em_htg_rod:*
                   36:  em_htg_mam:*
                   37:  em_htg_vrt:*
                   38:  em_sy:*
                   39:  em_hgo_hum:*
                   40:  em_hgo_mus:*
                   41:  em_hgo_other:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.		Score	% Query Match	Length	DB	ID	Description
1		517	100.0	294	6	AX431296	AX431296 Sequence
2		517	100.0	833	6	AR223870	AR223870 Sequence
3		517	100.0	1383	6	AX431306	AX431306 Sequence
4		517	100.0	1629	9	BC013316	BC013316 Homo sapi
5		517	100.0	2817	6	AX431318	AX431318 Sequence
6		517	100.0	2817	9	AF155118	AF155118 Homo sapi
7		517	100.0	2820	9	AK000528	AK000528 Homo sapi
8		513	99.2	1383	6	AX196260	AX196260 Sequence
9		513	99.2	1383	6	AF445802	AF445802 Homo sapi
10		476	92.1	1542	6	AX196262	AX196262 Sequence
11		476	92.1	1542	10	AF445803	AF445803 Homo sapi
12		476	92.1	2431	6	BC051676	BC051676 Mus muscu
13		413	79.9	501	6	AX321132	AX321132 Sequence
14		287	55.5	1719	5	BC045381	BC045381 Danio rer
15		256	49.5	33033	9	AY186092	AY186092 Homo sapi
16		256	49.5	118572	9	AC093012	AC093012 Homo sapi
17		256	49.5	165868	3	AC021719	AC021719 Homo sapi
18		256	49.5	305582	2	AC025567	AC025567 Homo sapi
19		256	49.5	211	6	AX431316	AX431316 Sequence
20		247	47.8	2213	9	BD155790	BD155790 Primer fo
21		247	47.8	2213	9	BD155790	BD155790 Primer fo
22		247	47.8	264657	2	AC129390	AC129390 Homo sapi
23		233	45.1	756	11	BV075713	BV075713 S212P6044
24		123	23.8	2069	3	AK116443	AK116443 Ciona int
25		121.5	23.5	1830	10	MMU440757	MMU440757 Mus muscu
26		98.5	19.1	1791	6	AX056430	AX056430 Sequence
27		96.5	18.7	2277	6	AX057324	AX057324 Sequence
28		96.5	18.7	2288	9	AF113136	AF113136 Homo sapi
29		96.5	18.7	1888	10	AF461763	AF461763 Mus muscu
30		95	18.4	161894	2	AC137339	AC137339 Homo sapi
31		88	17.0	249033	2	AC107528	AC107528 Rattus no
32		88	17.0	213254	2	AC144942	AC144942 Mus muscu
33		86.5	16.7	244696	2	AC123081	AC123081 Rattus no
34		86.5	16.7	252400	2	AC094859	AC094859 Rattus no
35		86	16.6	2603	5	BC043819	BC043819 Xenopus l
36		84	16.2	125777	9	AC078927	AC078927 Homo sapi
37		84	16.2	207951	2	AC012264	AC012264 Homo sapi
38		84	16.2	279125	2	AC113694	AC113694 Rattus no
39		84	16.2	288358	2	AC106121	AC106121 Rattus no
40		82	15.9	1947	10	BC005591	BC005591 Mus muscu
41		82	15.9	1960	10	BC051075	BC051075 Mus muscu
42		82	15.9	2606	10	MMU89023	MMU89023 Mus muscu
43		82	15.9	159431	2	AC128702	AC128702 Mus muscu
44		81	15.7	177210	2	AC134406	AC134406 Mus muscu
45		81	15.7	143452	2	AC136209	AC136209 Gallus ga
```
(c markers appear beside Result Nos. 8, 17, 18, 19, 23, 24, 34, 35, 36, 37, 38, 39, 40)
```

RESULT 1

ALIGNMENTS

LOCUS AX431296 294 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 5 from Patent WO0240680.
ACCESSION AX431296
VERSION AX431296.1 GI:21656165
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
Roth,W. and Stenner-Liewen,F.
TITLE Novel death domain proteins
JOURNAL Patent: WO 0240680-A 5 23-MAY-2002;
BURNHAM INST (US)
FEATURES Location/Qualifiers
source 1..294
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
CDS <1..>294
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD37274.1"
/db_xref="GI:21656166"
/translation="TYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSGDDRYNQFH
IRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLPDAVP"
BASE COUNT 84 a 55 c 62 g 93 t
ORIGIN

US-10-001-254-6 (1-98) x AX431296 (1-294)

Alignment Scores:
Pred. No.: 1.1le-61
Score: 517.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 6

Qy 1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
DB 1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
Qy 21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
DB 61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
Qy 41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
DB 121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
Qy 61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
DB 181 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 240
Qy 81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
DB 241 ATCCAAAATGAATTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCC 294

RESULT 2
AR223870
LOCUS AR223870 833 bp DNA linear PAT 26-SEP-2002
DEFINITION Sequence 10 from patent US 6440663.
ACCESSION AR223870
VERSION AR223870.1 GI:23332452
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 833)
AUTHORS Scanlan,M.J., Stockert,E., Chen,Y.-T., Old,L.J., Jager,E. and

---

TITLE Renal cancer associated antigens and uses therefor
JOURNAL Patent: US 6440663-A 10 27-AUG-2002;
Knuth,A.
FEATURES Location/Qualifiers
source 1..833
/organism="unknown"
BASE COUNT 273 a 155 c 179 g 226 t
ORIGIN

US-10-001-254-6 (1-98) x AR223870 (1-833)

Alignment Scores:
Pred. No.: 3.99e-61
Score: 517.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 6

Qy 1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
DB 74 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 133
Qy 21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
DB 134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 193
Qy 41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
DB 194 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 253
Qy 61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
DB 254 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 313
Qy 81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
DB 314 ATCCAAAATGAATTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCC 367

RESULT 3
AX431306
LOCUS AX431306 1383 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 15 from Patent WO0240680.
ACCESSION AX431306
VERSION AX431306.1 GI:21656175
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
Roth,W. and Stenner-Liewen,F.
TITLE Novel death domain proteins
JOURNAL Patent: WO 0240680-A 15 23-MAY-2002;
BURNHAM INST (US)
FEATURES Location/Qualifiers
source 1..1383
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
CDS 1..1383
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD37279.1"
/db_xref="GI:21656176"
/translation="MNKPITTPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
DDRYNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSTMPPDSSSPENSLEV
SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGPGVVYKGYVNNTTVAVKKLAAM
VDITTEELKQQFPQEIKVMAKCQHENLVELLGFSSSDGDDLCIVYVYMPNGSLLDRLSC
LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVLLEITTGLPAVDEHRE"

```
BASE COUNT      463 a    243 c    283 g    394 t
ORIGIN

        QOLILQEMTAS"
        POLILDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
```

Alignment Scores:

Pred. No.:
Score: 7.24e-61   Length: 1383
                                        Matches: 98
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 6   Indels: 0
DB: Gaps: 0

US-10-001-254-6 (1-98) x AX431306 (1-1383)

```
Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25 ACATATGTGCGCTGCCTCAATGTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 84

Qy  21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 144

Qy  41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 145 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 204

Qy  61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 205 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 264

Qy  81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
        |||||||||||||||||||||||||||||||||||||||||||||
Db 265 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 318
```

RESULT 4

```
LOCUS       BC013316               1629 bp    mRNA    linear   PRI 04-SEP-2001
DEFINITION  Homo sapiens, clone MGC:13330 IMAGE:4287014, mRNA, complete cds.
ACCESSION   BC013316
VERSION     BC013316.1  GI:15426431
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1629)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 19 Row: n Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7705840.
FEATURES
     source
            1..1629
            /organism="Homo sapiens"
```

```
     CDS
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="MGC:13330 IMAGE:4287014"
            /tissue_type="Brain, primitive neuroectodermal"
            /clone_lib="NIH_MGC_56"
            /lab_host="DH10B"
            /note="Vector: pDNR-LIB"
            /codon_start=1
            71..1453
            /product="unknown (protein for MGC:13330)"
            /protein_id="AAH13316.1"
            /db_xref="GI:15426432"
            /translation="MNKPITPSTYVRCLNVGLIIRKLSDFIDPQEGWKKLAVAIKKPSG
            DDRYNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
            DAVPKTANTLPSKEATVVQOKOMPFCDKDRTLMTPVQNLEQSYMPDSSSPENKSLEV
            SDTRFHSFSFYELKNVTNNPDERPISVGGNQMGEGGPGVVYKGYVNNTTVAVKKLAAM
            VDITTEELKQQFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
            LDGTPPLSWHRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAPTAKISDFGLAR
            ASEKPAQTVMTSRIVGTTAYMAPEALRGETPKSDIYSFGVVLLEIITGLPAVDEIRE
            POLILDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
            QOLILQEMTAS"
```

```
BASE COUNT      542 a    291 c    328 g    468 t
ORIGIN
```

US-10-001-254-6 (1-98) x BC013316 (1-1629)

Alignment Scores:

Pred. No.:
Score: 8.82e-61   Length: 1629
                                        Matches: 98
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 9   Indels: 0
DB: Gaps: 0

```
Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  95 ACATATGTGCGCTGCCTCAATGTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 154

Qy  21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 155 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 214

Qy  41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 215 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 274

Qy  61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 275 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 334

Qy  81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
        |||||||||||||||||||||||||||||||||||||||||||||
Db 335 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 388
```

RESULT 5
AX431318

```
LOCUS       AX431318               2817 bp    DNA     linear   PAT 28-JUN-2002
DEFINITION  Sequence 27 from Patent WO0240680.
ACCESSION   AX431318
VERSION     AX431318.1  GI:21656187
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
            Roth,W. and Stenner-Liewen,F.
  TITLE     Novel death domain proteins
  JOURNAL   Patent: WO 0240680-A 27 23-MAY-2002;
            BURNHAM INST (US)
FEATURES             Location/Qualifiers
     source
            1..2817
```

Left column

```
CDS
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                50..1432
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAD37285.1"
                /db_xref="GI:21656188"
                /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                DDRYNQFHIRRFEALLQTGKSPSTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                DAVPKTANTLPSKEAITVQQKOMPFCDKDRTLMTPVONLEQSYMPDSSSPENKSLEV
                SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                VDITTELKQOFDQEIKVMAKCQHENIVELLGFSSDGDDLCLVYVVMPNGSLLDRLSC
                LDGTPPLSWHMRCKIAQGAAANGINFLIEHNHIHRDIKSANILLDEAFTAKISDFGLAR
                ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                PQLLLDIKEEIEDEEKITEDYIDKKMNDADSTSVEAMYSGASQCRHEKKNKSPDIKKV
                HOLLQEMTAS"
BASE COUNT    912 a    547 c    586 g    772 t
ORIGIN
```

US-10-001-254-6 (1-98) x AX431318 (1-2817)

```
Alignment Scores:
Pred. No.:
Score:                  1.71e-60      Length:        2817
Percent Similarity:     517.00        Matches:       98
Best Local Similarity:  100.00%       Conservative:  0
Query Match:            100.00%       Mismatches:    0
                        6             Indels:        0
DB:                                   Gaps:          0
```

```
Qy    1   ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro  20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74   ACATATGTGCGCTGCCTCAATGTTGGACTAAATAGGAAGCTGTCAGATTTTATTGATCCT  133

Qy   21   GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr  40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134   CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC  193

Qy   41   AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer  60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  194   AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCAGTCT  253

Qy   61   GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  254   GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG  313

Qy   81   IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro  98
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  314   ATCCAAAATGAATTTTTTGCTCCCGCGAGTCTTTTGCTCCCAGATGCTGTTCCC  367
```

RESULT 6

```
AF155118
LOCUS       AF155118
DEFINITION  Homo sapiens putative protein kinase NY-REN-64 antigen mRNA,
            complete cds.
ACCESSION   AF155118
VERSION     AF155118.1  GI:5360130
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2817)
  AUTHORS   Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
            and Old,L.J.
```

```
  TITLE     Antigens recognized by autologous antibody in patients with
            renal-cell carcinoma
  JOURNAL   Int. J. Cancer 83 (4), 456-464 (1999)
  MEDLINE   99431124
   PUBMED   10508479
REFERENCE   2  (bases 1 to 2817)
  AUTHORS   Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
```

Right column

```
  AUTHORS   Jongeneel,V., Gure,A.O., Jager,D., Jager,E., Knuth,A., Chen,Y.-T.
            and Old,L.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Ludwig Institute, Sloan-Kettering
            Institute, 1275 York Ave, New York, NY 10021, USA
```

```
FEATURES             Location/Qualifiers
     source          1..2817
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="1973/10-4"
                     /cell_type="renal cell carcinoma"
     CDS             50..1432
                     /codon_start=1
                     /product="putative protein kinase NY-REN-64 antigen"
                     /protein_id="AAD42884.1"
                     /db_xref="GI:5360131"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPSTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQKOMPFCDKDRTLMTPVONLEQSYMPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTELKQOFDQEIKVMAKCQHENIVELLGFSSDGDDLCLVYVVMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAAANGINFLIEHNHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLLDIKEEIEDEEKITEDYIDKKMNDADSTSVEAMYSGASQCRHEKKNKSPDIKKV
                     HOLLQEMTAS"
BASE COUNT    912 a    547 c    586 g    772 t
ORIGIN
```

US-10-001-254-6 (1-98) x AF155118 (1-2817)

```
Alignment Scores:
Pred. No.:
Score:                  1.71e-60      Length:        2817
Percent Similarity:     517.00        Matches:       98
Best Local Similarity:  100.00%       Conservative:  0
Query Match:            100.00%       Mismatches:    0
                        9             Indels:        0
DB:                                   Gaps:          0
```

```
Qy    1   ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro  20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74   ACATATGTGCGCTGCCTCAATGTTGGACTAAATAGGAAGCTGTCAGATTTTATTGATCCT  133

Qy   21   GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr  40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134   CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC  193

Qy   41   AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer  60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  194   AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCAGTCT  253

Qy   61   GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  254   GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG  313

Qy   81   IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro  98
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  314   ATCCAAAATGAATTTTTTGCTCCCGCGAGTCTTTTGCTCCCAGATGCTGTTCCC  367
```

RESULT 7

```
AK000528
LOCUS       AK000528
DEFINITION  Homo sapiens cDNA FLJ20521 fis, clone KAT10395.
ACCESSION   AK000528
VERSION     AK000528.1  GI:7020683
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
            Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
```

Handwritten annotations: "2817 bp", "2820 bp", "DNA encoding SEQ ID #6"

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2820)

AUTHORS Nakamura,Y., Isogai,T. and Sugano,S.

AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.

TITLE Direct Submission

JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES Location/Qualifiers

source 1..2820
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KAT10395"
/cell_line="KATO III"
/cell_type="signet-ring cell carcinoma"
/clone_lib="KAT"
/note="cloning vector pME18SFL3".

CDS 30..1412
/codon_start=1
/note="unnamed protein product"
/protein_id="BAA91232.1"
/db_xref="GI:7020684"
/translation="MNKPITPSTYVRCLNVGLIRKLSDPIDPDEGWKKLAVAIKKPSG
DDRYNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
DAVPKTANTLPSKEAITVOQKOMPFCDKDRTLMTPVQNLEQSTMPPDSSSPENKSLEV
SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
VDITTEELKQOFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
ASEKPAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
POLLLDKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
QOLLQEMTAS"

BASE COUNT 534 a   579 c   767 g   940 t

ORIGIN

US-10-001-254-6 (1-98) x AK000528 (1-2820)

Alignment Scores:

Pred. No.:	
Score:	517.00	Length:	2820
Percent Similarity:	100.00%	Matches:	98
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	9	Mismatches:	0
DB:		Indels:	0
		Gaps:	0

Score: 1.71e-60

```
Qy      1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     54 ACATATGTGCGCTGCCTCAATGTTGGACTGATTAGGAAGCTGTCAGATTTTATTGATCCT  113

Qy     21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    114 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATAAAAAAACCATCTGGTGATGATAGATAC  173

Qy     41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    174 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT  233

Qy     61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    234 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGGATCTTTTG          293

Qy     81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro          98
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    234 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGGATCTTTTG           80
```

RESULT 8

Db 294 ATCCAAAATGAATTTTTTGCTCCTGCCAGTCTTTTGCTCCCCAGATGCTGTGTCCC 347

LOCUS AX196260   1383 bp   DNA   linear   PAT 28-AUG-2001

DEFINITION Sequence 2 from Patent WO0151641.

ACCESSION AX196260

VERSION AX196260.1 GI:15386462

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Wesche,H. and Li,S.

TITLE Irak-4: compositions and methods of use

JOURNAL Patent: WO 0151641-A 2 19-JUL-2001;
Tularik Inc. (US)

FEATURES Location/Qualifiers

source 1..1383
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="human Il-1 receptor-associated kinase 4 (IRAK-4) cDNA"

CDS 1..1383
/note="human IRAK-4"
/codon_start=1
/protein_id="CAC60090.1"
/db_xref="GI:15386463"
/translation="MNKPITPSTYVRCLNVGLIRKLSDPIDPDEGWKKLAVAIKKPSG
DDRYNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
DAVPKTANTLPSKEAITVOQKOMPFCDKDRTLMTPVQNLEQSTMPPDSSSPENKSLEV
SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
VDITTEELKQOFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
ASEKPAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
POLLLDKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
QOLLQEMTAS"

BASE COUNT 244 a   283 c   393 g   463 t

ORIGIN

US-10-001-254-6 (1-98) x AX196260 (1-1383)

Alignment Scores:

Pred. No.:	
Score:	513.00	Length:	1383
Percent Similarity:	98.98%	Matches:	97
Best Local Similarity:	98.98%	Conservative:	0
Query Match:	6	Mismatches:	1
DB:		Indels:	0
		Gaps:	0

Score: 2.56e-60

```
Qy      1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 ACATATGTGCGGTGCCTCAATGTTGGACTGATTAGGAAGCTGTCAGATTTTATTGATCCT   84

Qy     21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATAAAAAAACCATCTGGTGATGATAGATAC  144

Qy     41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT  204

Qy     61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGGATCTTTTG          264

Qy     81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro          98
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 ATCCAAAATGAATTTTTTGCTCCTCCTCCGAGTCTTTTGCTCCCCAGATGCTGTGTCCC  318
```

RESULT 9

```
LOCUS       AF445802              1383 bp    mRNA    linear   PRI 20-APR-2002
DEFINITION  Homo sapiens interleukin-1 receptor associated kinase 4 (IRAK4)
            mRNA, complete cds.
ACCESSION   AF445802
VERSION     AF445802.1  GI:20219009
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1383)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE     IRAK-4: a novel member of the IRAK family with the properties of an
            IRAK-kinase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (8), 5567-5572 (2002)
  MEDLINE   21957277
   PUBMED   11960013
REFERENCE   2  (bases 1 to 1383)
  AUTHORS   Suzuki,N., Suzuki,S., Duncan,G.S., Millar,D.G., Wada,T.,
            Mirtsos,C., Takada,H., Wakeham,A., Itie,A., Li,S., Penninger,J.M.,
            Wesche,H., Ohashi,P.S., Mak,T.W. and Yeh,W.C.
  TITLE     Severe impairment of interleukin-1 and Toll-like receptor
            signalling in mice lacking IRAK-4
  JOURNAL   Nature 416 (6882), 750-756 (2002)
  MEDLINE   21959395
   PUBMED   11923871
REFERENCE   3  (bases 1 to 1383)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) Biology I, Tularik Inc., 2 Corporate Drive,
            South San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1..1383
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /gene="IRAK4"
                     /db_xref="taxon:9606"
     gene            1..1383
                     /gene="IRAK4"
     CDS             1..1383
                     /gene="IRAK4"
                     /note="protein kinase"
                     /codon_start=1
                     /product="interleukin-1 receptor associated kinase 4"
                     /protein_id="AAM15772.1"
                     /db_xref="GI:20219010"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPSELLFDWGTTNCTPAGDLVDLLIQNFFAPASLLLP
                     DAVPKPANTLLSKEATIVTQVQQKDFKCDKDRTLMTPVDENQSYMPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTIVAVKKLAAM
                     VDITEELKQQFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIRDIKSANILLDKDTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEITTGLPAVDEHRE
                     QOLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNRPDIKKV"
BASE COUNT      463 a    244 c    283 g    393 t
ORIGIN
```

US-10-001-254-6 (1-98) x AF445802 (1-1383)

```
Alignment Scores:
Pred. No.:
Score:            2.56e-60        Length:        1383
                  513.00          Matches:       1383
Percent Similarity:   98.98%      Conservative:  97
Best Local Similarity: 98.98%     Mismatches:    1
Query Match:          99.23%      Indels:        0
                      9           Gaps:          0
```

```
Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACATATGTCCGGCTCCTCAATGTTGGACTTATTAGGAAGCTGTCAGATTTTATTGATCCT 84

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaAlaIleLysLysProSerGlyAspAspArgTyr 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGAGAC 144
```

```
RESULT 10
LOCUS       AX196262              1542 bp    DNA     linear   PAT 28-AUG-2001
DEFINITION  Sequence 4 from Patent WO0151641.
ACCESSION   AX196262
VERSION     AX196262.1  GI:15386464
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1542)
  AUTHORS   Wesche,H. and Li,S.
  TITLE     Irak-4: compositions and methods of use
  JOURNAL   Patent: WO 0151641-A 4 19-JUL-2001;
            Tularik Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1542
                     /organism="Mus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10095"
                     /note="murine Il-1 receptor-associated kinase 4;
                     cDNA"
     CDS             163..1542
                     /note="murine IRAK-4"
                     /codon_start=1
                     /protein_id="CAC60091.1"
                     /db_xref="GI:15386465"
                     /translation="MNKPLTPSTYIRNLNVGILRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLLIQEELFAPATLLLP
                     DAVPQTVKSLPREAATVAQTHGPCQEKDRTSVMPKLEHSCEPPDSSSPDNRSVES
                     SDTRFHSFSFHELKSITNNFDEQPASAGGNRMGEGGFGVVYKGVVNNTIVAVKKLGAM
                     VEISTEELKQQFDQEIKVMATCQHENLVELLGFSSDSDNLCLVYAYMPNGSLLDRLSC
                     LDGTPPLSWHARCKVAQGTANGIRFLHENHHIRDIKSANILLDKDTAKISDFGLAR
                     ASARLAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEITTGLAAVDENRE
                     QOLLDIKEEIEDYDEKMSDADPASVEAMYSAASQCLHEKKNRPDIAKV"
BASE COUNT      421 a    392 c    423 g    306 t
ORIGIN
```

US-10-001-254-6 (1-98) x AX196262 (1-1542)

```
Alignment Scores:
Pred. No.:
Score:            3.5e-55         Length:        1542
                  476.00          Matches:       89
Percent Similarity:   95.92%      Conservative:  5
Best Local Similarity: 90.82%     Mismatches:    4
Query Match:          92.07%      Indels:        0
                      6           Gaps:          0
```

```
Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
          |||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db  187 ACATACATACGGCAACCTTAATGTGGGGATCCTTAGGAAGCTGTCGGATTTTATTGATCCT 246

Qy   21 GlnGluGlyTrpLysLysLeuSerLeuAlaValAlaAlaIleLysLysProSerGlyAspAspArgTyr 40
          |||||||||||||||||:::||:::|||||:::|||||||:::|||||||||||||||||
Db  247 CAAGAAGGGTGGAAGAAAATTAGCAGTAGCTATCAAAAAGCCGTCCGGCGACGACAGATAC 306
```

```
Qy    41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
                                                                         .
Db   307 AATCAGTTCCATATAAGGAGGATTCGAAAGCCTTACTTCAGACGGGAGGAAGCCCCACCTGT 366

Qy    61 GluLeuLeuPheAspThrProGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 GAACTGCTGTTTGACTGGGGGACCACAGAACTGCACAGTTGGCGACCTTTGTGGATCTACTG 426

Qy    81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
            |||:::||
Db   427 GTCCAGATTGAGCTGTTTGCCCCGGCCACTCTTCTGCTTCCGGATGCCGTTCC 480
```

RESULT 11

```
LOCUS          AF445803                1542 bp    mRNA    linear    ROD 20-APR-2002
DEFINITION     Mus musculus interleukin-1 receptor associated kinase 4 (Irak4)
               mRNA, complete cds.
ACCESSION      AF445803
VERSION        AF445803.1  GI:20219011
KEYWORDS       .
SOURCE         Mus musculus (house mouse)
  ORGANISM     Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1  (bases 1 to 1542)
  AUTHORS      Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE        IRAK-4: a novel member of the IRAK family with the properties of an
               IRAK-kinase
  JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 99 (8), 5567-5572 (2002)
  MEDLINE      21957277
   PUBMED      11960013
REFERENCE      2  (bases 1 to 1542)
  AUTHORS      Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE        Direct Submission
  JOURNAL      Submitted (08-NOV-2001) Biology I, Tularik Inc., 2 Corporate Drive,
               South San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1..1542
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
     gene            1..1542
                     /gene="Irak4"
     CDS             163..1542
                     /gene="Irak4"
                     /note="protein kinase"
                     /codon_start=1
                     /product="interleukin-1 receptor associated kinase 4"
                     /protein_id="AAM15773.1"
                     /db_xref="GI:20219012"
                     /translation="MNKLTPSTYIRNLNVGILRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLLVQIELFAPATLLLP
                     DAVPQTVKSLPPRRAATVAQTHGPCQEKDRTSVMPRKLEHSCEPPDSSSPDNRSVES
                     VEISTEELKQOFPQEIKVMATCQHENIVELLGFSSDSDNLCLVYAYMPNGSLLDRLSC
                     LDGTPPLSMHTRCKVAQGTANGIRFLHENHHIHRDIKSANILLDKDFTAKISDFGLAR
                     ASARLAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLELITGLAAVDENRE
                     PQLLDIKEEIEDEEKTIEDYTDEKMSDADPASVEAMYSAASQCLHEKKNRRPDIAKV
                     QQLLQEMSA"
BASE COUNT      421 a      392 c      423 g      306 t
ORIGIN
```

US-10-001-254-6 (1-98) x AF445803 (1-1542)

Alignment Scores:
```
Pred. No.:                       3.5e-55         Length:           1542
Score:                          476.00           Matches:            89
Percent Similarity:              95.92%          Conservative:        5
Best Local Similarity:           90.82%          Mismatches:          4
Query Match:                     92.07%          Indels:              0
DB:                              10              Gaps:                0
```

```
Qy     1 ThrTyrValArgCysLeuAsnValGlyIleLeuArgLysLeuSerAspPheIleAspPro 20
           |:::||  ||||||||||    |||:::|||||:|||||||||||||:|||||||:||
Db   187 ACATACACGACGCAACCTTAAICTGGGGATCCTTAGGAAGCTTAGTGATTTTATTGATCCT 246

Qy    21 GlnGluGlyTrpLysLysLeuAlaValAlaAlaAlaGlyLysLysProSerGlyAspAspArgTyr 40
           ||||||||||||||||||||||||||    ||||||||||||||||||||||||||||||
Db   247 CAAGAAGGGTGGAAGAAATTAGCAGTTGCTATCAAAAAGCCGTCCGGAGACGACAGATAC 306

Qy    41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrGlyLysSerProThrSer 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307 AATCAGTTCCATATAAGGAGGATTCGAAAGCCTTACTTCAGACGGGAGGAAGCCCCACCTGT 366

Qy    61 GluLeuLeuPheAspThrProGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 GAACTGCTGTTTGACTGGGGGACCACAGAACTGCACAGTTGGCGACCTTTGTGGATCTACTG 426

Qy    81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
           |||:::||
Db   427 GTCCAGATTGAGCTGTTTGCCCCGGCCACTCTTCTGCTTCCGGATGCCGTTCC 480
```

RESULT 12

```
LOCUS          BC051676                2431 bp    mRNA    linear    ROD 10-JUN-2003
DEFINITION     Mus musculus interleukin-1 receptor-associated kinase 4, transcript
               variant 1, mRNA (cDNA clone MGC:60994 IMAGE:30017484), complete
               cds.
ACCESSION      BC051676
VERSION        BC051676.1  GI:30354527
KEYWORDS       MGC.
SOURCE         Mus musculus (house mouse)
  ORGANISM     Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1  (bases 1 to 2431)
  AUTHORS      Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
               Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
               Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
               Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
               Diatchenko,L., Matsumura,K., Farmer,A.A., Rubin,G.M., Hong,L.,
               Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
               Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
               Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
               Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
               McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
               Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
               Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
               Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodriguez,S.,
               Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
               Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
               Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
               Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
               Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE        Generation and initial analysis of more than 15,000 full-length
               human and mouse cDNA sequences
  JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE      22388257
   PUBMED      12477932
REFERENCE      2  (bases 1 to 2431)
  AUTHORS      Strausberg,R.
  TITLE        Direct Submission
  JOURNAL      Submitted (30-APR-2003) National Institutes of Health, Mammalian
               Gene Collection (MGC), Cancer Genomics Office, National Cancer
               Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
               USA
REMARK         NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT        Contact: MGC help desk
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Minoru Ko
               cDNA Library Preparation: Yulan Piao and Minoru Ko
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Baylor College of Medicine Human Genome
               Sequencing Center
               Center code: BCM-HGSC
```

Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

This clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 111 Row: 1 Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 23943897.

```
FEATURES             Location/Qualifiers
     source          1..2431
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="MGC:60994 IMAGE:30017484"
                     /tissue_type="Embryo, whole, mouse, 7.5-dpc"
                     /clone_lib="NIA Mouse 7.5-dpc Whole Embryo cDNA Library
                     (long)"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1"
     gene            1..2431
                     /gene="Irak4"
                     /note="synonyms: NY-REN-64, 9330209D03Rik"
                     /db_xref="LocusID:266632"
                     /db_xref="MGI:2182474"
     CDS             168..1547
                     /codon_start=1
                     /product="interleukin-1 receptor-associated kinase 4,
                     isoform a"
                     /protein_id="AAH51676.1"
                     /db_xref="GI:30545528"
                     /db_xref="LocusID:266632"
                     /db_xref="MGI:2182474"
                     /translation="MNKPLTPSTYIRNILNVGILRKLSDFIPDQEGWKKLAVAIKKPSG
                     DDRYNQPHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLLVQIELFAPATLLLP
                     DAVPQTVKSLIPREAATVAQTHGPCQEKDRTSVMPMPKLEHSCEPFDSSSPDNRSVES
                     SDTRFHSFSFHELKSITNNFDEQPASAGNRMGEGGFGVVYKGCVNNTIVAVKKLGAM
                     VEISTEELKQQFDQEIKVMATCQHENLVELLGFSSDSDNLCLVYAYMPNGSLLDRLSC
                     LDGTPPLSWHTRCKVAQGTANGIRFLHENHIHRDIKSANILLDRDFTAKISDFGLAR
                     ASARLAQTVMISRIVGTTAYMAPEALRGEITFKSDIYSFGVVLLELITGLAAVDENRE
                     PQLLLDIKEEIEDEEKIIEDYIDEKMSDADPASVEAMYSASAQCLHEKKNRRPDIAKV
                     QQLLLQEMSA"
```

BASE COUNT   656 a   580 c   618 g   577 t
ORIGIN

US-10-001-254-6 (1-98) x BC051676 (1-2431)

Alignment Scores:
Pred. No.:
Score: 6.0e-55   Length: 2431
Percent Similarity: 95.92%   Matches: 89
Best Local Similarity: 92.08%   Conservative: 5
Query Match: 10   Mismatches: 4
                                  Indels: 0
                                  Gaps: 0

```
Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
        ||| ||:|||||||||:::|||||||||||||||||||||||||||||||||||||||
Db  192 ACATACATACGGACAACCTTAANTGGGGAATCCTTAGGAAGCTGTCCGGATTTTATTGATCCT 251

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  252 CAAGAAGGGTGGGAAGAAAATTAGCAGTAGCTATCAAAAAGCCGTCCGGCGACGACAGATAC 311

Qy   41 AsnGlnPheHisIleArgArgPheGluLeuLeuLeuGlnThrGlyLysSerProThrSer 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  312 AATCAGTTCCATATAAGGAGGTTCGAAGCCTTACTTCAGACCGGGAAGAGCCCACCTCT 371

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  372 GAACTGCTGTTTGACTGGGGCACCACCAACTGCACAGTTGGCGACCTTGTGGATCTACTG 431
```

Qy   81 IleGlnAsnGluPheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
        ||| ||| |||:::|||||||||||||||||||||||||||||||||||||||
Db 432 GTCCAGATTGAGCTGTTTGCCCCGCCGCCACTCTCCCTGCCTGACCGGATGCGTTCCC 485

RESULT 13
```
LOCUS       AX321132
DEFINITION  Sequence 149 from Patent WO0177168.
ACCESSION   AX321132
VERSION     AX321132.1  GI:17904576
KEYWORDS
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Lodes,M.J., Wang,T., Mohamath,R. and Indrias,C.Y.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0177168-A 149 18-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1..501
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```
Db AX321132 Sequence 149 from Patent WO0177168.   501 bp   DNA   linear   PAT 15-DEC-2001

BASE COUNT   166 a   102 c   97 g   134 t   2 others
ORIGIN

US-10-001-254-6 (1-98) x AX321132 (1-501)

Alignment Scores:
Pred. No.:
Score: 4e-47   Length: 501
Percent Similarity: 94.00%   Matches: 93
Best Local Similarity: 93.00%   Conservative: 1
Query Match: 6   Mismatches: 4
                                  Indels: 4
                                  Gaps: 0

```
Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   45 ACATATGTGCGGCCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 104

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105 CAAGAAGGATGGAAGAAGTTAGCTGTATTTAAAAAACCATCTGGTGATGATAGAGATAC 164

Qy   41 AsnGln-PheHisIleArgArgPheGluLeuLeuLeuGlnThrGlyLysSerProThrSe 60
        ||||| ||||||||||||||||||||||||||||||||||||  |||||||||||||||
Db  165 AATCAAGTTCACATAAGGAGAGATTTGAAGCAATT-CTTCAAACTGGAAAAAGTCCACTTC 223

Qy   60 rGluLeuLeuPheAspTrp-GlyThrThrAsnCysThrValGlyAspLeuValAspLeuL 80
        :||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  224 TTGAATACTCGTTGACTGGGGGGCCACCACAAATTGGACACAGTTGGTGATCTTGTGGATCTT 283

Qy   80 euIleGlnAsnGluPheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  284 TGATCCAAAATGAATT-TTTGCTCCTGCGAGTCTTTTGCTCCCCAGATGCGTGTTCCC 338
```

RESULT 14
```
LOCUS       BC045381                1719 bp    mRNA    linear   VRT 10-FEB-2003
DEFINITION  Danio rerio, Similar to interleukin-1 receptor-associated kinase 4,
            clone MGC:55553 IMAGE:2642773, mRNA, complete cds.
ACCESSION   BC045381
VERSION     BC045381.1  GI:28278875
KEYWORDS    MGC.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 1719)
```

```
AUTHORS    Strausberg,R.
TITLE      Direct Submission
JOURNAL    Submitted (31-JAN-2003) National Institutes of Health, Mammalian
           Gene Collection (MGC), Cancer Genomics Office, National Cancer
           Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
           USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
           cDNA Library Preparation: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Institute for Systems Biology
           http://www.systemsbiology.org
           contact: amadan@systemsbiology.org
           Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
           Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
           Clone distribution: MGC clone distribution information can be found
           through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
           Series: IRAK Plate: 101 Row: m Column: 22
           This clone was selected for full length sequencing because it
           passed the following selection criteria: Hexamer frequency ORF
           analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..1719
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="MGC:55553 IMAGE:2642773"
                     /tissue_type="Whole body, adult, (one male and one female,
                     including unfertilized eggs)"
                     /clone_lib="Sugano Kawakami_zebrafish_DRA"
                     /lab_host="DH10B"
                     /note="Vector: pME18S-FL3"
     CDS             122..1270
                     /codon_start=1
                     /product="Similar to interleukin-1 receptor-associated
                     kinase 4"
                     /protein_id="AAH45381.1"
                     /db_xref="GI:28278876"
                     /translation="MSDVTPNTPVRKLRYSALRALADLLDPQDTWRSIMADISRPCGE
                     PRYTQMHMRFEACVLQGKSPTMELLFPWGTSPCTVGDLVEIIIFNIQLFAAVTVLLPD
                     HSVCHAHTGSVWCEEAAPASAVCLQACEITQTVEDNSNKPQKISKPVEDDSNKPVQEQ
                     ELFVEPDSSSGAQESSWDSSQGFHTFSLHELTAMTQHWDERPLSDGCRLGSGGFGVV
                     FRGRMGDKHVAVKKLNPLDGSSYEDLRKQFNQEIQTLRSLSHSHNVLRVLGCSCSGPPL
                     CVVFELMVNGSLLERLACAEHTPALTWRNRCWITVGAARGLSYLHTHAHIHRDVKSAN
                     ILLDEGFVAKISDFGLTRSAAGAAGGFVWASSGRESGSCTVAGDEG"
BASE COUNT     492 a   399 c   477 g   351 t
ORIGIN
```

```
Qy          61 GluLeuLeuPheAspThrGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         323 GAGCTGCTGTTCGACACTGGGGGAAACCTCAGACTGCCACTGTTGGAGATCTTGTTGGAGATCCTG 382

Qy          81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuLeuProAsp 95
               ||||||||||||  |||||  |||||||| ||||||||||||||||||
Db         383 ATCCGGCCATCAGCTGTGTGCGCGCCGTCACAGTCCTGCTGCCCGAT 427
```

US-10-001-254-6 (1-98) x BC045381 (1-1719)

```
Alignment Scores:
Pred. No.:                           Length:      1719
Score:            3.5e-29            Matches:       51
Score:            287.00            Conservative:   19
Percent Similarity:  73.68%         Mismatches:     25
Best Local Similarity: 53.68%       Indels:          0
Query Match.:         55.51%        Gaps:            0
DB:                   5
```

```
Qy           1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuSerAspPheIleAspPro 20
               |||||||||||| ||| ||  ||||||||  |||| ||||||||| |||
Db         143 ACACCGGGTTCGAAACCTCCGTACTCCGCCGCTAAGAGCTTTAGCAGACCTGCTGGACCCG 202

Qy          21 GlnGluGlyTrpLysLeuValAlaValAlaIleIleTyrGlyTyrProSerGlyTyrAspAspArgTyr 40
               |||| ||  |  |  ||   ::  :  :   :  ||||  :: |  :
Db         203 CAGGACACCTGGAGGAGCATCATGGCGGGACATCAGCAGACCTGCGGGGGAACCCAGATAC 262

Qy          41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
               |||  |  |:||  |||||||||| ||||||||||||||||||||||
Db         263 ACACAGATGCACATGAGGCGCGGTTTGAAGCATGCAGTCCTGCAGGGGAAAGAGTCCCACCATG 322
```

```
LOCUS       AY186092               33033 bp    DNA    linear   PRI 02-DEC-2002
```

```
FEATURES             Location/Qualifiers
     source          1..33033
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     repeat_region   34..88
                     /rpt_family="L2"
     repeat_region   89..383
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       92..94
                     /replace=""
     variation       133
                     /frequency="0.02"
                     /replace="a"
     repeat_region   384..599
                     /rpt_family="L2"
                     /rpt_type=dispersed
     variation       642
                     /frequency="0.02"
                     /replace="c"
     variation       708..941
                     /rpt_family="L2"
                     /rpt_type=dispersed
     repeat_region   854
                     /frequency="0.02"
                     /replace="a"
     variation       1265
                     /frequency="0.01"
                     /replace="c"
     variation       1533..1534
                     /frequency="0.93"
                     /replace=""
     gene            1894..31033
                     /gene="IRAK4"
     mRNA            join(1894..1933,11040..11209,14157..14302,15117..15299,
                     15849..16000,16902..16966,20567..20681,21130..21239,
                     25244..25427,26599..26661,29336..29494,29617..31033)
                     /gene="IRAK4"
                     /product="interleukin-1 receptor-associated kinase 4"
     variation       2315
                     /gene="IRAK4"
```

variation   1265
                /frequency="0.06"
                /replace="a"

variation   642
                /frequency="0.02"
                /rpt_type=dispersed

repeat_region   384..599
                /rpt_family="L2"
                /rpt_type=dispersed

```
variation       2365
                /frequency="0.01"
                /gene="IRAK4"
                /replace="c"

variation       2456
                /gene="IRAK4"
                /frequency="0.01"
                /replace="t"

variation       2506. .2807
                /gene="IRAK4"
                /frequency="0.09"
                /replace="t"

repeat_region   2940
                /rpt_family="Alu"
                /rpt_type=dispersed

variation       3023
                /gene="IRAK4"
                /frequency="0.04"
                /replace="t"

variation       3025. .3915
                /gene="IRAK4"
                /frequency="0.72"
                /replace="t"

repeat_region   3038
                /rpt_family="L2"
                /rpt_type=dispersed

variation       3115. .3129
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"

variation       3189
                /gene="IRAK4"
                /frequency="0.01"
                /replace="t"

variation       3536
                /gene="IRAK4"
                /frequency="0.14"
                /replace="c"

variation       3573
                /gene="IRAK4"
                /frequency="0.01"
                /replace="t"

variation       3819
                /gene="IRAK4"
                /frequency="0.17"
                /replace="t"

variation       3877
                /gene="IRAK4"
                /frequency="0.01"
                /replace="g"

variation       4144
                /gene="IRAK4"
                /frequency="0.04"
                /replace="a"

repeat_region   4438. .4610
                /rpt_family="MIR"
                /rpt_type=dispersed

variation       4602
                /gene="IRAK4"
                /frequency="0.01"
                /replace=""

variation       4797
                /gene="IRAK4"
                /frequency="0.01"
                /replace="t"

variation       4972
                /gene="IRAK4"
                /frequency="0.02"
                /replace="a"

repeat_region   5759. .5797
                /rpt_family="L2"
                /rpt_type=dispersed

variation       5821
                /gene="IRAK4"
                /frequency="0.02"
                /replace="g"

variation       5880
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"

variation       5954
                /gene="IRAK4"
                /frequency="0.01"
                /replace="t"

repeat_region   5964. .6036
                /rpt_family="MER1_type"
                /rpt_type=dispersed

repeat_region   6023. .6590
                /rpt_family="MER1_type"
                /rpt_type=dispersed

variation       6204
                /gene="IRAK4"
                /frequency="0.04"
                /replace="t"

variation       6661. .6662
                /gene="IRAK4"
                /frequency="0.04"
                /replace=""

variation       6998
                /gene="IRAK4"
                /frequency="0.04"
                /replace=""

repeat_region   7015. .7073
                /rpt_family="Mariner"
                /rpt_type=dispersed

repeat_region   7074. .7366
                /rpt_family="Alu"
                /rpt_type=dispersed

variation       7122
                /gene="IRAK4"
                /frequency="0.01"
                /replace="a"

variation       7147
                /gene="IRAK4"
                /frequency="0.01"
                /replace="a"

variation       7234
                /gene="IRAK4"
                /frequency="0.04"
                /replace="c"

repeat_region   7367. .8636
                /rpt_family="Mariner"
                /rpt_type=dispersed

variation       7392
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"

variation       7597
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"

variation       7794
                /gene="IRAK4"
                /frequency="0.09"
                /replace="g"
```

Alignment Scores:

Pred. No.:	2.5e-23	Length:	33033
Score:	256.00	Matches:	49
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	98.00%	Mismatches:	0
Query Match:	49.52%	Indels:	0
DB:	9	Gaps:	0

```
US-10-001-254-6 (1-98) x AY186092 (1-33033)

Qy       45 IleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPhe 64
            ::: ||||||||||||||||||||||||||||||||||||||||||||||||
Db    14152 TTAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTT 14211

Qy       65 AspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsnGlu 84
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14212 GACTGGGGCACCACCAAATTGCACAGTTGGTGTGATCTTGTGGATCTTTTGATCCAAAATGAA 14271

Qy       85 PhePheAlaProAlaSerLeuLeuLeuPro 94
            ||||||||||||||||||||||||||||||
Db    14272 TTTTTTGCTCCTGCGAGTCTTTTGCTCCCA 14301
```

Search completed: January 18, 2004, 04:26:13
Job time : 1667.38 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 00:40:22 ; Search time 117,536 Seconds
                   (without alignments)
                   2250.748 Million cell updates/sec

Title: US-10-001-254-6
Perfect score: 517
Sequence: 1 TYVRCLNVGLIRKLSDFIDP..........LLIONEFFAPASLLPDAVP 98

Scoring table: BLOSUM62
                   Xgapop 10.0 , Xgapext   0.5
                   Ygapop 10.0 , Ygapext   0.5
                   Fgapop 6.0 , Fgapext   7.0
                   Delop   6.0 , Delext   7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO spool/US10001254/runat 16012004_152423_19704/app_query.fasta_1.1109
-DB=N-Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOFCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10001254 @CGN 1_1_0 @runat 16012004_152423_19704 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseqn-embl/NA2003.DAT:*

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	517	100.0	294	24	AAD40074	Human IRAK4 DD (de
2	517	100.0	833	21	AAA09319	Human cancer assoc
3.	517	100.0	1383	24	AAD40079	Human IRAK4 gene #
4	517	100.0	1668	23	AAS76805	DNA encoding novel
5	517	100.0	2817	24	AAD40085	Human IRAK4 gene #
6	513	99.1	1383	21	AAD10197	Human interleukin-
7	476	92.1	1542	22	AAD10198	Mouse interleukin-
8	413	79.9	501	24	AAS61608	Lung small cell ca
9	256	49.5	405	23	AAS76803	DNA encoding novel
10	247	47.8	211	24	AAD40084	Human IRAK4 short
11	247	47.8	415	23	AAS76802	DNA encoding novel
12	247	47.8	2213	22	AAH13798	Human cDNA sequenc
13	134.5	26.0	408	24	ABQ56733	Human colon cancer
14	96.5	18.7	1791	22	AAF44695	Novel protein kina
15	96.5	18.7	2277	21	AAC84459	Interleukin-1 rece
16	96.5	18.7	2288	21	AAZ61602	DNA encoding an in
17	96.5	18.7	2293	22	AAI57800	Human polynucleoti
18	96.5	18.7	2490	23	AAS86902	DNA encoding novel
19	84	16.2	1494	23	AAS67194	DNA encoding novel
20	79	15.3	1806	20	AAX77599	Human IRAK-2alpha
21	79	15.3	3459	20	AAX77600	Human IRAK-2beta c
22	78.5	15.2	861	22	AAL37224	Human musculoskele
23	78.5	15.2	861	25	AEX60212	cDNA encoding nove
24	78.5	15.2	864	22	ABX60210	Human musculoskele
25	78.5	15.2	864	21	AAL37225	Human musculoskele
26	78.5	15.2	864	25	ABX60210	cDNA encoding nove
27	78.5	15.2	864	25	ABX60013	cDNA encoding nove
28	74.5	14.4	652	24	ABK53369	Human eosinophil-m
29	74.5	14.4	774	-24	ABK73720	Bacillus lichenifo
30	72.5	14.0	2819	24	ABK52386	DNA encoding cysti
31	72.5	14.0	2828	24	ABQ91977	Human NF-kB activa
32	72.5	14.0	2828	24	ABK52388	DNA encoding cysti
33	72.5	14.0	2829	21	AAC75235	Human ORFX ORF790
34	72.5	14.0	3191	23	ABV23079	Human prostate exp
35	72.5	14.0	3191	23	ABV28915	Human prostate exp
36	72	13.9	2344	22	ABL16597	Drosophila melanog
37	72	13.9	4574	23	ABL16596	Drosophila melanog
38	72	13.9	534720	19	AAV30458	cDNA encoding nove
39	72	13.9	531165	19	AAV30459	cDNA encoding nove
40	71.5	13.8	527	22	ABA43336	Human breast cance
41	70.5	13.6	470	22	AAO20039	Human breast cell
42	70.5	13.6	470	22	AAK27496	Human brain expres
43	70.5	13.6	470	22	AAI33409	Human bone marrow
44	70.5	13.6	470	22	AAI01995	Probe #2095 used t
45	70.5	13.6	470	22	Probe #1986 used t

ALIGNMENTS

RESULT 1
AAD40074
ID   AAD40074 standard; cDNA; 294 BP.
XX
AC   AAD40074;
XX
DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 DD (death domain) cDNA.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ss.
OS   Homo sapiens.

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

```
XX
FH   Key              Location/Qualifiers
FT   CDS              1..294
FT                    /*tag= a
FT                    /product= "Human IRAK4 DD"
FT                    /note= "No start and stop codon"
FT                    /partial
XX
PN   WO200240680-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US044844.
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24854.
XX
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
PA   (BURN-) BURNHAM INST.
XX
PT   29-JUN-2001; 2001US-301889P.
PR   17-NOV-2000; 2000US-0715893.
XX
PR   allergy, sepsis, autoimmunity, allograft rejection and other diseases
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   New polypeptide comprising a death domain or death effector domain,
XX
PS   Claim 18; Page 173-174; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, by contacting a DD, DED or
CC   that binds a DD, DED or NB-ARC domain. The invention is useful for
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis , C.
CC   muridarum, C. pneumoniae; and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 DD cDNA.
XX
SQ   Sequence 294 BP; 84 A; 55 C; 62 G; 93 T; 0 other;
XX
Alignment Scores:
Pred. No.:       3.94e-64
Score:           517.00
Percent Similarity:   100.00%
Best Local Similarity:  100.00%
Query Match:     24

Length:      294
Matches:     294
Conservative:   0
Mismatches:     0
Indels:      0
Gaps:        0
```

```
US-10-001-254-6 (1-98) x AAD40074 (1-294)

Qy   1 ThrTyrValArgCysLeuAsnValGlyTyrLeuIleArgGlyLeuSerAspPheIleAspPro 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1 ACATATGTCCGCTGCCTCAATGTTGGACTAATGTTAGGAAGCGTGTCAGATTTTATTGATCCT 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   21 GlnGlyGlyTrpLysLysLeuAlaValAlaIleAlaIleLysLysProSerGlyAspAspArgTyr 40
Db   61 CAAGGAGGATGGAAGAAGTTAGCTGTAGCTATAGCAATAAAACCATCTGGTGATGATAGATAC 120
Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuGlnThrGlyLysSerProThrSer 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AATCAGTTTCACATAAGGAGGTTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysArgThrValGlyAspLeuValAspLeuLeu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGACCTTTTG 240
Qy   81 IleGlnAsnGlyPheHisLeuArgGluSerAspPheIleAspPro 98
        ||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ATCCAAAATGAATTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGTCTGTTCCC 294
```

```
RESULT 2
AAA09319
ID   AAA09319 standard; DNA, 833 BP.
XX
AC   AAA09319;
XX
DT   10-AUG-2000  (first entry)
XX
FH   Key              Location/Qualifiers
FT   CDS              50..670
FT                    /*tag= a
XX
DE   Human cancer associated antigen precursor DNA, clone NY-REN-64.
XX
OS   Homo sapiens.
XX
KW   renal cancer; cancer associated antigen precursor; diagnosis;
KW   cytostatic; ss.
XX
PN   WO200020587-A2.
XX
PD   13-APR-2000.
XX
PF   04-OCT-1999; 99WO-US22873.
XX
PR   05-OCT-1998; 98US-0166300.
PR   05-OCT-1998; 98US-0166350.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y, Gout I, Tureci O, Sahin U, Pfreundschuh M, Scanlan MJ;
PI   Stockert E, Chen Y, Old LJ, Jager E, Knuth A;
XX
DR   WPI; 2000-303774/26.
DR   P-PSDB; AAY92347.
XX
PS   Claim 57; Page 85; 121pp; English.
XX
CC   AAA09310-20 are novel genes isolated by SEREX screening from a renal
CC   cancer cell line 1973/10.4. The genes encode cancer associated antigen
CC   precursors. These gene products are useful in methods for preventing,
CC   diagnosing and/or treating disorders, especially cancer, associated with
CC   abnormal expression of human cancer associated antigens. The method
CC   comprises contacting a sample from a subject with an agent that
CC   specifically binds to the nucleic acid molecule or expression product
CC   (or fragment) complexed with a human leukocyte antigen (HLA) molecule
CC   and determining the interaction between the agent and the nucleic acid
CC   molecule or the expression product as a determination of the disorder.
XX
SQ   Sequence 833 BP; 273 A; 155 C; 179 G; 226 T; 0 other;
XX
Alignment Scores:
```

Pred. No.:   1.68e-63   Length: 833
Score: 517.00   Matches: 98
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: Gaps: 0
Query Match: 21

US-10-001-254-6 (1-98) x AAA09319 (1-833)

```
Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  74 ACATATGTGCGCGTGCCTCAATGTGGACTAATTAGGAGGCTGTCAGATTTTATTGATCCT 133

Qy  21 GlnGluGlyTrpLysLeuLeuAlaValAlaIleLeuLysProSerGlyAspAspArgTyr 40
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134 CAAGAAGGATGGAAGAAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 193

Qy  41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrGlyLysSerProThrSer 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 194 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCACTTCT 253

Qy  61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254 GAATTACTGTTTGACTGGGGCACCACAAAATTGCCACAGTTGGTGGATCTTTTG 313

Qy  81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 314 ATCCAAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCC 367
```

RESULT 3

ID   AAD40079 standard; DNA; 1383 BP.
XX
AC   AAD40079;
XX
DT   22-OCT-2002   (first entry)
XX
DE   Human IRAK4 gene #1.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1383
FT   /*tag= a
FT   /product= "Human IRAK4"
XX
PN   WO200240680-A2.
PD   23-MAY-2002.
PF   15-NOV-2001; 2001WO-US44844.
PR   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.
PA   (BURN-) BURNHAM INST.
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24859.
XX
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection and other diseases

XX
PS   Claim 19; Page 180-182; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis, C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 gene.
XX
SQ   Sequence 1383 BP; 463 A; 243 C; 283 G; 394 T; 0 other;

Alignment Scores:
Pred. No.:   3.4e-63   Length: 1383
Score: 517.00   Matches: 98
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: Gaps: 0
Query Match: 24

US-10-001-254-6 (1-98) x AAD40079 (1-1383)

```
Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25 ACATATGTGCGCGTGCCTCAATGTGGACTAATTAGGAGGCTGTCAGATTTTATTGATCCT 84

Qy  21 GlnGluGlyTrpLysLeuLeuAlaValAlaIleLeuLysProSerGlyAspAspArgTyr 40
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  85 CAAGAAGGATGGAAGAAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 144

Qy  41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrGlyLysSerProThrSer 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 145 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCACTTCT 204

Qy  61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 205 GAATTACTGTTTGACTGGGGCACCACAAAATTGCCACAGTTGGTGGATCTTTTG 264

Qy  81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 265 ATCCAAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCC 318
```

RESULT 4

ID   AAS76805 standard; cDNA; 1668 BP.
XX
AC   AAS76805;
XX
DT   13-FEB-2002   (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12609.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12618.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 12609; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1668 BP; 571 A; 289 C; 332 G; 476 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 4.41e-63
Score: 517.00
Percent Similarity: 100.00%   Length: 1668
Best Local Similarity: 100.00%   Matches: 1668   98
Query Match: 100.00%   Conservative: 0
Best Local Match: Mismatches: 0
DB: 23   Indels: 0
                                       Gaps: 0

US-10-001-254-6 (1-98) x AAS76805 (1-1668)

Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT   84

Qy   21 GlnLeuGlyTrpLysGlyLeuAlaValAlaIleLeuValGlySerGlyAspAspArgTyr   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAAGAAGGATGGAAGGGAAGAGTTAGCTGTAGCTATAAAAAAACCATCTGGTGTGATGATGATAC 144

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 145 AATCAGTTTCACATAAGGAGGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 204

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValValAspLeuLeu 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 205 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGTGGATCTTTG 264

Qy   81 IleGlnAsnGluPheAlaProAlaSerLeuLeuProAspAlaValPro 98
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db 265 ATCCAAAATGAATTTTTGCTCCTCGAGTCTTTTGCCTCCAGATGCTGTTCCC 318

RESULT 5
AAD40085
ID   AAD40085 standard; DNA; 2817 BP.
XX
AC   AAD40085;
XX
DE   Human IRAK4 gene #2.
XX
DT   22-OCT-2002 (first entry)
XX
PN   WO200240680-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US44844.
XX
PR   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24865.
XX
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection and other diseases
PT   -
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   50..1432
FT   /*tag= a
FT   /product= "Human IRAK4"
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
XX
PS   Claim 19; Page 194-196; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis, C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for

Column 1

CC treating bacterial infections by modulating the activity of bacterial
CC proteins. The protein and antibody specific for it are useful for
CC discovery of drugs that suppress infection, inflammation, allergy,
CC sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC is useful for treating immune-based pathologies, pathologies associated
CC with cell division, inflammatory diseases such as sepsis, fibrosis,
CC arthritis, graft versus host disease. The invention is used in antisense
CC therapy and gene therapy. The present sequence is human IRAK4 gene.
XX
SQ Sequence 2817 BP; 912 A; 547 C; 586 G; 772 T; 0 other;

US-10-001-254-6 (1-98) x AAD40085 (1-2817)

Alignment Scores:
Pred. No.:
Score: 9.15e-63
Percent Similarity: 517.00   Length: 2817
Best Local Similarity: 100.00%   Matches: 98
Query Match: 100.00%   Conservative: 0
DB: 24   Mismatches: 0
                                        Indels: 0
                                        Gaps: 0

Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgGlySLeuSerAspPheIleAspPro 20
Db   74 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAGGGCTGTCTCAGATTTATTGATCCT 133

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysGlySerGlyAspAspArgTyr 40
Db 134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAGCCATCTGGTGATGATAGATAC 193

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
Db 194 AATCAGTTTCACATAAGGAGGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCACTTCT 253

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
Db 254 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGGATCTTGTGGATCTTTTG 313

Qy   81 IleGlnAsnGluPheGluAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
Db 314 ATCCAAAATGAATTTTTGCTCCTCGAGTCTTTTGCTCCAGATGCTGTGTTCCC 367

RESULT 6
ID   AAD10197 standard; cDNA; 1383 BP.
XX
AC   AAD10197;
XX
DT   24-SEP-2001 (first entry)
XX
DE   Human interleukin-1 receptor-associated kinase (IRAK)-4 cDNA.
XX
KW   Human; interleukin-1 receptor associated kinase-4; IRAK-4; cytostatic;
KW   IL; antibacterial; antiinflammatory; ophthalmological; vascotropic; OPD;
KW   chronic obstructive pulmonary disease; neuroprotective; chronic cough;
KW   adult respiratory disease syndrome; pulmonary fibrosis; asthma; ARDS;
KW   interstitial lung disease; allergic rhinitis; transplant rejection;
KW   autoimmune disease; rheumatoid arthritis; systemic lupus erythematosus;
KW   multiple sclerosis; diabetes; cancer; solid tumour; lymphoma; stroke;
KW   cardiovascular disease; atherosclerosis; neurodegenerative disease;
KW   sepsis; osteoarthritis; rash; osteoporosis; psoriasis; dermatitis;
KW   inflammatory bowel disease; gout; Crohn's disease; ulcerative colitis;
KW   Behcet's syndrome; ankylosing spondylitis; IL-1 receptor/Toll receptor;
KW   sarcoidosis; transgenic animal; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1383
FT   /*tag= a
FT   /product= "Human IRAK-4"
XX
PN   WO200151641-A1.

Column 2

PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US01171.
XX
PR   13-JAN-2000; 2000US-0176395.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Wesche H, Li S;
XX
DR   WPI; 2001-451860/48.
DR   P-PSDB; AAB05398.
XX.
PT   Novel human interleukin-1 receptor associated kinase polypeptide,
PT   useful for identifying modulators of the polypeptide for treating gout,
PT   asthma, allergic rhinitis, multiple sclerosis and skin cancer -
XX
PS   Claim 7; Fig 2; 89pp; English.
XX
CC   The present sequence is a cDNA encoding human interleukin (IL)-1 receptor
CC   associated kinase (IRAK)-4. IRAK associate with activated IL-1, IL-18
CC   and other receptors and act to transduce signals originating from the
CC   activated receptors, ultimately leading to a variety of downstream
CC   effects such as nuclear factor (NF)-kappaB activation. The IRAK-4
CC   inhibitors are useful for treating inflammatory diseases such as
CC   pulmonary diseases and diseases of the airway (e.g., adult respiratory
CC   disease syndrome (ARDS), chronic obstructive pulmonary disease (OPD),
CC   pulmonary fibrosis, interstitial lung disease, asthma, chronic cough
CC   or allergic rhinitis), transplant rejection, autoimmune diseases (e.g.,
CC   rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis
CC   or diabetes), cancer (e.g., solid tumours, skin cancer or lymphoma),
CC   cardiovascular diseases (e.g., stroke and atherosclerosis), diseases
CC   of the central nervous system (e.g., neurodegenerative disease), CD14
CC   mediated sepsis, non-CD14 mediated sepsis, osteoarthritis, osteoporosis,
CC   psoriasis, diseases of the skin (e.g., rash, contact dermatitis, atopic
CC   dermatitis), inflammatory bowel disease (e.g., Crohn's disease and
CC   ulcerative colitis), Behcet's syndrome, ankylosing spondylitis, gout,
CC   sarcoidosis and ophthalmic diseases and conditions. The inhibitors of
CC   IRAK-4 activity or expression are used to inhibit signal transduction
CC   resulting from the activation of an interleukin-1 receptor (IL-1R)/Toll
CC   receptor in a cell. They also inhibit the activation of a transcription
CC   factor that activates NFkappaB in the cell. IRAK-4 is used to create a
CC   nonhuman transgenic animal which is useful for testing the function of
CC   IRAK-4 in vivo, to generate models for the study of potential treatments
CC   for IRAK-4-related inflammatory diseases and conditions. IRAK-4 sequences
CC   disorders and conditions and for the development of potential treatments
CC   are also used in gene therapy and in antisense therapy.
XX
SQ   Sequence 1383 BP; 463 A; 244 C; 283 G; 393 T; 0 other;

US-10-001-254-6 (1-98) x AAD10197 (1-1383)

Alignment Scores:
Pred. No.:
Score: 1.26e-62
Percent Similarity: 513.00   Length: 1383
Best Local Similarity: 98.98%   Matches: 97
Query Match: 99.23%   Conservative: 0
DB: 22   Mismatches: 1
                                        Indels: 0
                                        Gaps: 0

Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgGlySLeuSerAspPheIleAspPro 20
Db   25 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAGGGCTGTCTCAGATTTATTGATCCT 84

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysGlySerGlyAspAspArgTyr 40
Db   85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAGCCATCTGGTGATGATAGATAC 144

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
Db 145 AATCAGTTTCACATAAGGAGGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCACTTCT 204

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80

```
Db
       205  GAATTACTGTTGACTCGGGCACCACAAATTGACAGCTGGTGGATCTTGTTG  264
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      81  IleGlnAsnGluPhePheAlaProAlaSerLeuLeuProAspAlaValPro  98

Db     265  ATCCAAAATGAATTTTTGCTCCTGCGAGATCTTTTGCTCCAGATTGCTGTTCCC  318
            |||||||||||||||||||||||||||||||||||||||||||
```

RESULT 7

AAD10198

DE   Mouse interleukin-1 (IL-1) receptor-associated kinase (IRAK)-4 cDNA.
XX
DT   24-SEP-2001 (first entry)
XX
AC   AAD10198;
XX
ID   AAD10198 standard; cDNA; 1542 BP.
XX

KW   Mouse; interleukin-1 receptor associated kinase-4; IRAK-4; cytostatic;
KW   IL; antibacterial; antiinflammatory; ophthalmological; vasotropic; OPD;
KW   chronic obstructive pulmonary disease; neuroprotective; chronic cough;
KW   adult respiratory disease syndrome; pulmonary fibrosis; asthma; ARDS;
KW   interstitial lung disease; allergic rhinitis; transplant rejection;
KW   autoimmune disease; rheumatoid arthritis; systemic lupus erythematosus;
KW   multiple sclerosis; diabetes; cancer; solid tumour; lymphoma; stroke;
KW   cardiovascular disease; atherosclerosis; neurodegenerative disease;
KW   sepsis; osteoarthritis; rash; osteoporosis; psoriasis; dermatitis;
KW   inflammatory bowel disease; gout; Crohn's disease; ulcerative colitis;
KW   Behcet's syndrome; ankylosing spondylitis; IL-1 receptor/Toll receptor;
KW   sarcoidosis; transgenic animal; ss.
XX
OS   Mus sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1542
FT   /*tag= a
FT   /product= "Mouse IRAK-4"
XX
PN   WO200151641-A1.
XX
PD   19-JUL-2001.
XX
Qy   12-JAN-2001; 2001WO-US01171.
XX
PR   13-JAN-2000; 2000US-0176395.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Weeche H, Li S;
XX
DR   WPI; 2001-451860/48.
DR   P-PSDB; AAE05399.
XX
PS   Claim 26; Fig 4; 89pp; English.
XX

CC   The present sequence is a cDNA encoding mouse interleukin (IL)-1 receptor
CC   associated kinase (IRAK)-4. IRAK associate with activated IL-1, IL-18
CC   and other receptors and act to transduce signals originating from the
CC   activated receptors, ultimately leading to a variety of downstream
CC   effects such as nuclear factor (NF)-kappaB activation. The IRAK-4
CC   inhibitors are useful for treating inflammatory diseases such as
CC   pulmonary diseases and diseases of the airway (e.g., adult respiratory
CC   disease syndrome (ARDS), chronic obstructive pulmonary disease (OPD),
CC   pulmonary fibrosis, interstitial lung disease, asthma, chronic cough
CC   or allergic rhinitis), systemic lupus erythematosus, multiple sclerosis
CC   rheumatoid arthritis, transplant rejection, autoimmune diseases (e.g.,
CC   or diabetes), cancer (e.g. solid tumours, skin cancer or lymphoma),
CC   cardiovascular diseases (e.g., stroke and atherosclerosis), diseases
CC   of the central nervous system (e.g.; neurodegenerative disease), CD14
CC   mediated sepsis, non-CD14 mediated sepsis, osteoarthritis, osteoporosis,
```

---

Right column:

```
CC    psoriasis, diseases of the skin (e.g., rash), contact dermatitis, atopic
CC    dermatitis), inflammatory bowel disease (e.g., Crohn's disease and
CC    ulcerative colitis), Behcet's syndrome, ankylosing spondylitis, gout,
CC    sarcoidosis and ophthalmic diseases and conditions. The inhibitors of
CC    IRAK-4 activity or expression are used to inhibit signal transduction
CC    resulting from the activation of an interleukin-1 receptor (IL-1R)/Toll
CC    receptor in a cell. They also inhibit the activation of a transcription
CC    factor that activates NFkappaB in the cell. IRAK-4 is used to create a
CC    nonhuman transgenic animal which is useful for testing the function of
CC    IRAK-4 in vivo, to generate models for the study of inflammatory
CC    disorders and conditions and for the development of potential treatments
CC    for IRAK-4-related inflammatory diseases and conditions. IRAK-4 sequences
CC    are also used in gene therapy and in antisense therapy.
XX
SQ    Sequence 1542 BP; 421 A; 392 C; 423 G; 306 T; 0 other;

Alignment Scores:
Pred. No.:              .
Score:                  476.00         Length:          1542
Percent Similarity:     95.92%         Matches:           89
Best Local Similarity:  90.82%         Conservative:       5
Query Match:            92.07%         Mismatches:         4
DB:                     22             Indels:             0
                                       Gaps:               0

US-10-001-254-6 (1-98) x AAD10198 (1-1542)

Qy      1  ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro  20
           |||||||||||||||::||||||||||||||||||||||||||||||||||||||||||
Db      1  ACATTACATACGCAACCTTAATGTGGGATCCTTAGGAAGCTGTCCGGATTTTATTGATCCT  60

Qy     21  GlnGlyTrpLysGlyLeuAlaValAlaIleLysLysLysProSerGlyAspAspArgTyr  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  CAAGAAAGGGTGGAAGAAAATTAGCAGTTAGCTATTCAAAAAAGCCGTCCGGCGACGACAGATAC  306

Qy     41  AsnGlnPheHisIleGluArgArgPheGluAlaLeuLeuGlnThrHrGlyLysSerProThrSer  60
           ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187  ACAGGAAGGGGTGGAAGAAAATTAGCAGTTAGCTATTCAAAAAAGCCGTCCGGCGACGACAGATAC  306

Qy     61  GluLeuLeuProAspAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspAspLeuLeu  80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247  GAACTGCTGTTTGACTGGGGCACCACGAACTGCACAGTTGGGGACTTGGTCGGATTTTATTGATCCT  426

Qy     81  IleGlnAsnGluPhePheAlaProAlaSerLeuLeuProAspAlaValPro  98
           :::|||             |||    ||||||||||||||||||||
Db    427  GTCCAGATTGAGCTGTTTGCCCCCCCCCACTCCTCCTGTCGCCGGATGCCGTTCCC  480

RESULT 8

AAS61608

DE    Lung small cell carcinoma antigen, cDNA #149.
XX
DT    29-JAN-2002  (first entry)
XX
AC    AAS61608;
XX
ID    AAS61608 standard; cDNA; 501 BP.
XX
KW    Human; cytostatic; antitumour; lung small cell cancer antigen;
KW    tumour; lung cancer; ss.
XX
OS    Homo sapiens.
XX
PN    WO200171168-A2.
XX
PD    18-OCT-2001.
XX
PF    11-APR-2001; 2001WO-US11859.
XX
PR    11-APR-2000; 2000US-196780P.
PR    21-JUN-2000; 2000US-213361P.
PR    01-SEP-2000; 2000US-227763P.
PR    05-SEP-2000; 2000US-230629P.
PR    14-SEP-2000; 2000US-232565P.
PR    19-DEC-2000; 2000US-257037P.
```

Left column

```
PR   08-JAN-2001; 2001US-260796P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Lodes MJ, Wang T, Mohamath R, Indirias CY;
XX
DR   WPI; 2002-010896/01.
XX
PT   Lung tumour polynucleotide and polypeptides useful in therapy and
PT   diagnosis of cancer especially lung cancer -
XX
PS   Claim 1; Page 174; 295pp; English.
XX
CC   The invention relates to novel isolated lung small cell cancer antigen
CC   polynucleotides (I) and polypeptides (II) used in a method of detecting
CC   cancer in a patient. The method is optionally performed by
CC   utilising oligonucleotides (III), where the biological sample
CC   from the patient is contacted with (III), detecting the amount of
CC   polynucleotide hybridised to (III) in the sample and comparing the
CC   amount of polynucleotide to a predetermined cut-off value and thereby
CC   determining cancer in a patient. (I), (II) or antigen-presenting cells
CC   expressing (II) is useful for stimulating T cells
CC   specific for a tumour protein. The method comprises contacting T cells
CC   with one of the components under conditions to permit the stimulation
CC   and/or expansion of the cells. A composition comprising (I) is useful for
CC   stimulating an immune response in a patient and for inhibiting the
CC   development of a cancer especially lung cancer in a patient. An
CC   isolated T cell population is useful for removing tumour cells from the
CC   biological sample and for inhibiting the development of cancer in a
CC   patient. AAS61460-AAS61874 represent novel human lung small cell
CC   cancer antigen coding sequences of the invention.
XX
SQ   Sequence 501 BP; 166 A; 102 C; 97 G; 134 T; 2 other;
```

```
US-10-001-254-6 (1-98) x AAS61608 (1-501)

Alignment Scores:
Pred. No.:              4.94e-49
Score:                  413.00
Percent Similarity:     94.00%       Length:        501
Best Local Similarity:  93.00%       Matches:       93
Query Match:            79.88%       Conservative:  1
                        24           Mismatches:    4
                                     Indels:        4
                                     Gaps:          0

Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   45 ACATAAGTGGCCTGCCTCAATGTTGGACTAATTAGGAAGCGTGTCAGATTTTATTGATTCCT 104

Qy   21 GlnGluGlyTrpLysLysLeuValAlaValAlaIleLeuLysLysProSerGlyAspAspArgTyr 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGAGTAGATAC 164

Qy   41 AsnGln-PheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSe 60
        |||||||   |||||||||||||||||||||||||||||||||||||||||
Db  165 AATCAAGTTTCACATAAGGAGAGATTTGAAGCATT-CTTCAAACTGGAAAAAGTCCCACTTC 223

Qy   61 rGluLeuLeuPheAspTrp-GlyThrThrAsnCysThrValGlyAspLeuValAspLeuL 80
        ||  ::::|||||||||||| |||||||||||||||||||||||||||||||||
Db  224 TTGAATACTGTTTGACTGGGGCGGCCACCACAAATTGGACAGTTGGGTGGATCTTTGTGGA 283

Qy   81 euIleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuPheAspAlaValPro 98
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  284 TGATCCAAAATGAATT-TTTGCTCCTGCGAGTTCTTTGCTCCTCCAGATGCTGTTCCC 338
```

```
RESULT 9
ID   AAS76803 standard; cDNA; 405 BP.
XX
AC   AAS76803;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12607.
```

Right column

```
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12616.
XX
PR   30-MAR-2001; 2001WO-US08631.
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PN   11-OCT-2001.
XX
PS   Claim 1; SEQ ID No 12607; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and encoded polypeptides, useful in
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (III). (I) and (III) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations in
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 405 BP; 93 A; 87 C; 83 G; 142 T; 0 other;
```

```
US-10-001-254-6 (1-98) x AAS76803 (1-405)

Alignment Scores:
Pred. No.:              7.44e-27
Score:                  256.00
Percent Similarity:     100.00%      Length:        405
Best Local Similarity:  98.00%       Matches:       49
Query Match:            49.52%       Conservative:  1
                        23           Mismatches:    0
                                     Indels:        0
                                     Gaps:          0

Qy   45 IleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPhe 64
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  114 TTAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTT 173

Qy   65 AspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuValIleGlnAsnGlu 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174 GACTGGGGCGGCCACCACAAATTGGACAGTTGGGTGGATCTTTGTGGATCAAAATGAA 233

Qy   85 PhePheAlaProAlaSerLeuLeuLeuPhe 94
        ||||||||||||||||||||||||||||||
```

Db   234 TTTTTGCTCCTCGAGTCTTTGCTCCCA 263

RESULT 10

ID   AAD40084 standard; DNA; 211 BP.
XX
AC   AAD40084;
XX
DT   22-OCT-2002   (first entry)
XX
DE   Human IRAK4 short gene.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..180
FT   /tag= a
FT   /product= "Human IRAK4 short protein"
XX
PN   WO200240680-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US44844.
XX
PR   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24864.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection, and other diseases
PT   -
XX
PS   Claim 19; Page 195; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis , C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,

CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 short gene.
XX
SQ   Sequence 211 BP; 77 A; 41 C; 36 G; 57 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 5.7e-26
Percent Similarity: 247.00   Length: 211
Best Local Similarity: 100.00%   Matches: 46
Query Match: 100.00%   Conservative: 0
DB: 47.78%   Mismatches: 0
                            24   Indels: 0
                                            Gaps: 0

Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgGlyLysLeuSerAspPheIleAspPro 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAGGAAGCTGTCAGATTTTATTGATCCT 20

Qy   21 GlnGluGlyTrpLysGlyLysLeuAlaValAlaIleLysLysProSerGlyTyrAspAspArgTyr 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAGGAAGCTGTCAGATTTTATTGATCCT 84

Qy   41 AsnGlnPheHisIleArg 46
               |||||||||||||||||
Db   85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGAGATTAC 144

Db   145 AATCAGTTTCACAATAAGA 162

RESULT 11

ID   AAS76802 standard; cDNA; 415 BP.
XX
AC   AAS76802;
XX
DT   13-FEB-2002   (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12606.
XX
OS   Homo sapiens.
XX
PN   WO20017SO67-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSB-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12615.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 12606; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 415 BP; 144 A; 83 C; 78 G; 110 T; 0 other;

US-10-001-254-6 (1-98) x AAS76802 (1-415)

Alignment Scores:
Pred. No.:   1.46e-25
Score: 247.00
Percent Similarity: 100.00%
Percent Similarity: 100.00%
Best Local Similarity: 47.78%
Query Match: 23

Length: 415
Matches: 46
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

Qy 1 ThrThrTyrValArgCysLeuAsnValGlyLeuIleLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 60 ACATAATGTGCGCCTGCCTCAATGTTGGACTAATTAGGAAGAGCGTGTCAGATTTTATTGATCCT 119

Qy 21 GlnGluGlyTrpProLysLysLeuValAlaValAlaLeuAlaIleLeuLysLysProSerGlyAspAspAspArgTyr 40
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 120 CAAGAAGGATGGGAAGAAGTTAGCTGTTAGCTAATTAAAAAAACCATTCTGGTGATGATAGTAGATAC 179

Qy 41 AsnGlnPheHisIleArg 46
       ||||||||||||||||||
Db 180 AATCAGTTTCACATAAGA 197

RESULT 12
AAH13798
ID AAH13798 standard; cDNA; 2213 BP.
XX
AC AAH13798;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:10742.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
XX
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-016126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0018776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 10742; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs easily without any specialised methods.
CC AAH13633 to AAH18742 represent human cDNA sequences; AAA92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2213 BP; 642 A; 488 C; 516 G; 567 T; 0 other;

US-10-001-254-6 (1-98) x AAH13798 (1-2213)

Alignment Scores:
Pred. No.:   1.5e-24
Score: 247.00
Percent Similarity: 100.00%
Best Local Similarity: 47.78%
Query Match: 22

Length: 2213
Matches: 46
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

Qy 1 ThrThrTyrValArgCysLeuAsnValGlyLeuIleLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 775 ACATAATGTGCGCCTGCCTCAATGTTGGACTAATTAGGAAGAGCGTGTCAGATTTTATTGATCCT 834

Qy 21 GlnGluGlyTrpProLysLysLeuValAlaValAlaLeuAlaIleLeuLysLysProSerGlyAspAspAspArgTyr 40
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 835 CAAGAAGGATGGGAAGAAGTTAGCTGTTAGCTAATTAAAAAAACCATTCTGGTGATGATAGTAGATAC 894

Qy 41 AsnGlnPheHisIleArg 46
       ||||||||||||||||||
Db 895 AATCAGTTTCACATAAGA 912

RESULT 13
ABQ56733
ID ABQ56733 standard; cDNA; 408 BP.
XX
AC ABQ56733;
XX
DT 02-AUG-2002 (first entry)
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:428.
XX
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW genetic analysis; diagnostic; antisense therapy; gene; ss.
OS Homo sapiens.
XX
XX
PN WO200229086-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30732.
XX

PR
PA
XX
PA   (FARB) BAYER CORP.
XX
PI   Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI   Thiaglingam A, Lewis ME;
XX
DR
XX
WPI; 2002-426115/45.
XX
PS
XX
PT   New isolated nucleic acid that is differentially expressed in cancer
PT   tissues useful for determining the presence of colon cancer in a cell
PT   or tissue type, and in antisense therapy -
XX
PS
XX
Claim 1; Fig 1; 796pp; English.
XX
CC   ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC   expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
CC   encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC   used in antisense therapy. An antibody immunoreactive with a polypeptide
CC   encoded by (I) is useful for detecting cancer in a patient sample, and
CC   for detecting the presence or absence of a polynucleotide encoded by a
CC   nucleic acid which hybridies to (I) in a cell. A probe/primer derived
CC   from (I) can be used for determining the presence of a nucleic acid which
CC   hybridies to (I), and for determining the phenotype of cells in a sample
CC   of cells from a patient. (I) is useful for determining the presence of
CC   colon cancer in a cell or tissue type, for determining the presence or
CC   state of other type of cancer, in antisense therapy, to generate
CC   macroarrays on a solid surface, to identify a chromosome on which the
CC   corresponding gene resides, and in tissue profiling, forensics, genetic
CC   analysis, mapping and diagnostic applications. (I) can be used to raise
CC   antibodies, and to screen for peptide analogues and antagonists.
XX
SQ   Sequence 408 BP; 114 A; 92 C; 81 G; 61 T; 60 other;

US-10-001-254-6 (1-98) x ABQ56733 (1-408)

Alignment Scores:
Pred. No.:   1.38e-09
Score: 134.50
Percent Similarity: 52.17%
Best Local Similarity: 50.00%
Query Match: 26.02%

Length: 408
Matches: 46
Conservative: 2
Mismatches: 38
Indels: 7
Gaps: 1

```
Qy      9 GlyLeuIleArgLysLeuSerAspPheIleAspProGlnGlyGlyTrpLysLysLeuAla  28
            |||||||||||||||||||||||   |||||||||| |||||||||||||||  |||||
Db     63 GGNCTAATNANGAAGCNGNCAGANTTTANCGANCCTNNANANGANG-GANAAGGCAGCT 121

Qy     29 ValAlaIleLysLysProSerGlyAspAsp-ArgTyrAsnGlnPheHis-IleArgArgP  48
            ||||||||||||||||||||  ||||||||  ||||||||||| ||  ||||||||||
Db    122 GNACCTATNAAAAAACCANCNGGGGATGANNGATACCATCAGTTTCACANANNGGAGAT 181

Qy     48 heGluAlaLeuLeu-GlnThrGly---LysSerProThrSerGluLeuLeuPheAspTrp  66
            | | |||||||||| ::|||||||   ||| |||||||||||||||| | |||||| |
Db    182 NTGAAGCAttaCTNNAAAACTGGGAAAAAAGGCCCNCTTCGGAANGAACTGGTTGACTGG 241

Qy     67 GlyThrThrAsnCysThr-ValGlyAspLeuValAspLeuLeuIleGlnAsn-GluPheP  86
            ||||||||||||||||| ||||||||||||||||||||||||||||||||| |||||||
Db    242 GGCACCACAAAATTCCACAGNTGGGGGGATCTTGTGNATNCNCGAANCACAACNGAATTTN 301

Qy     86 heAlaProAlaSerLeuLeuLeuPro  94
            || ||| |||||||||||||||||
Db    302 TNGCTCCNTGCNANGCCNGGGCNCC 327
```

```
Qy       68  ThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
             |||:|||::|:||||||:||||||::: ::|||
Db      232  CAGAAAAACAAGACCATCGGTGACCTTTTACAGGTCCTC 270
```

```
Qy        8  ValGlyLeuIleArgLysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeu 27
             ::::||||                       :::   ::: |||    ||:::  |||
Db       74  CTCGGAGAGCTCTCGCCTGTTCTCGGACAGCTGCCACACGCGCGCTGGGGTGGGCGCCTG 133
```

```
Qy       28  AlaValAlaIleLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArg 47
             |||                       --------------         :::|||
Db      134  GCAGAGAGACTTTCAAGC-------------------------AGCTGGCTGGATGTTCGTCAT 172
```

```
Qy       48  PheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGly 67
             |||                       |||||||||       |||||||||:::
Db      173  ATTGAAAAAGTATGTGAGACCAAGGTAAAAAGTGGAACAAGAGAATTACTTTGGTCCTGGGGCA 232
```

```
Qy       68  ThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
             |||:|||::|:||||||:||||||::: ::|||
Db      233  CAGAAAAACAAGACCATCGGTGACCTTTTACAGGTCCTC 271
```

Search completed: January 18, 2004, 01:31:59
Job time : 120.536 secs

```
Db      232  CAGAAAAACAAGACCATCGGTGACCTTTTACAGGTCCTC 270
Qy       68  ThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
             |||:|||::|:||||||:||||||::: ::|||
```

AC   AAC84459;
XX
ID   AAC84459 standard; cDNA; 2277 BP.
XX
AC   AAC84459;
XX
DT   02-APR-2001 (first entry)
XX
DE   Interleukin-1 receptor associated kinase-3 (IRAK3) encoding cDNA.
XX
KW   IRAK3; interleukin-1 receptor associated kinase-3; drug screening;
KW   interleukin-1; IL-1; antiinflammatory; cytostatic; immunomodulator;
KW   immunosuppressive; vulnerary; anti-human immunodeficiency virus; HIV;
KW   nootropic; neuroprotective; antiarthritic; antiasthmatic; respiratory;
KW   osteopathic; vasotropic; antiarteriosclerotic; cerebroprotective; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   2..1792
FT   /*tag= a
FT   /transl_except= (pos:656..658, aa: Ile)
XX
PN   WO20007531O-A1.
XX
PD   14-DEC-2000.
XX
PF   08-JUN-2000; 2000WO-EP05290.
XX
PR   08-JUN-1999;   99GB-0013180.
XX
PA   (GLAX) GLAXO GROUP LTD.
XX
PI   Volpe F;
XX
DR   WPI; 2001-071072/08.
DR   P-PSDB; AAB48173.
XX
PT   Interleukin-1 receptor associated kinases, useful for screening drugs
PT   for treating diseases affected by interleukin-1 activity, e.g.
PT   inflammations or autoimmune diseases -
XX
PS   Claim 12; Page 22-24; 26pp; English.
XX
CC   The invention provides interleukin-1 receptor associated kinase-3 (IRAK3)
CC   proteins. The IRAK3 proteins are useful in drug screening, particularly
CC   for screening modulators of IRAK3. The modulator is useful for treating
CC   a disease or condition associated with interleukin-1 (IL-1) family
CC   activity. It may also be used for manufacturing a medicament for the
CC   treatment of these diseases or conditions which include inflammations,
CC   post-operative septicemia autoimmune diseases (e.g. Crohn's disease,
CC   psoriasis, multiple sclerosis, transplant rejection or graft versus
CC   host disease), stroke, ischemia, reperfusion injury, acute respiratory
CC   disease syndrome, restenosis, atherosclerosis, AIDS (acquired immune
CC   deficiency syndrome), bone diseases (e.g. osteoporosis), cancer,
CC   Alzheimer's disease, rheumatoid arthritis, poor wound healing or
CC   asthma. The present sequence represents an IRAK3 protein encoding cDNA.
XX
SQ   Sequence 2277 BP; 709 A; 469 C; 501 G; 598 T; 0 other;
```

Alignment Scores:
Pred. No.:
Score: 0.00377
Percent Similarity: 96.50
Best Local Similarity: 47.95%
Query Match: 31.51%
DB: 18.67%
                                                           22

Length: 2277
Matches: 23
Conservative: 12
Mismatches: 31
Indels: 7
Gaps: 1

US-10-001-254-6 (1-98) x AAC84459 (1-2277)

US-10-001-254-6 (1-98) x AAC84459 (1-2277)

```
OM protein - nucleic search, using frame_plus_p2n model

Run on: January 18, 2004, 04:27:15

                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2004 Compugen Ltd.

Title:          US-10-001-254-6
Perfect score:  517
Sequence:       1 TVVRCLNVGLIRKLSDFIDP............LLIQNEFPAPASLLPDAVP 98

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop  6.0 , Fgapext 7.0
                Delop   6.0 , Delext  7.0

Searched:       2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters:   4648192

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters :
 -MODEL=frame+_p2n.model -DEV=xlh
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2.1/USPTO_spool/US10001254/runat_16012004_152426_19815/app_query.fasta_1.1109
 -DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
 -LOOPCL=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
 -LOOPEXT=0 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
 -TRANS=human40.cdi -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
 -MAXLEN=200000000 -USER=US10001254 @CGN 1_1_534 @runat_16012004_152426_19815
 -NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
       Published_Applications_NA:*
 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Search time 146,603 Seconds (without alignments)
2356.207 Million cell updates/sec

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	517	100.0	294	15	US-10-001-254-5	Sequence 5, Appli
2	517	100.0	1383	15	US-10-001-254-15	Sequence 15, Appli
3	517	100.0	2817	15	US-09-966-451-3	Sequence 3, Appli
4	517	100.0	2817	15	US-10-001-254-27	Sequence 27, Appli
5	513	99.2	1383	11	US-09-759-595-2	Sequence 1, Appli
6	476	92.1	1542	11	US-09-759-595-4	Sequence 4, Appli
7	413	79.9	501	9	US-09-833-790-149	Sequence 149, App
8	256	49.5	31000	11	US-09-966-451-11	Sequence 10, App
9	256	49.5	211	15	US-10-001-254-25	Sequence 25, Appl
10	247	47.8	2288	9	US-09-863-549-1	Sequence 1, Appli
11	96.5	18.7	2293	14	US-09-773-753-1	Sequence 1, Appli
12	95	18.4	1888	13	US-10-340-545-1	Sequence 1, Appli
13	79.5	15.4	476	13	US-10-366-288-19	Sequence 19, Appl
14	79	15.3	1782	13	US-09-764-877-3590	Sequence 3590, A
15	79	15.3	3128	12	US-10-108-260A-709	Sequence 709, App
16	79	15.3	1806	9	US-09-773-753-3	Sequence 3, Appli
17	78.5	15.2	3459	9	US-09-974-300-1011	Sequence 1011, A
18	78.5	15.2	479	10	US-09-773-753-14	Sequence 14, Appl
19	78.5	15.2	861	10	US-09-764-877-3589	Sequence 3589, A
20	78.5	15.2	864	10	US-09-764-877-3587	Sequence 3587, Ap
21	78.5	14.6	864	10	US-09-764-877-3590	Sequence 3590, A
22	75.5	14.6	3128	12	US-10-108-260A-709	Sequence 709, App
23	74.5	14.4	774	10	US-09-974-300-1011	Sequence 1011, A
24	73	14.1	497	13	US-10-027-632-50723	Sequence 50723, A
25	73	14.1	497	14	US-10-042-298A-132	Sequence 132, App
26	73	14.1	499	13	US-10-027-632-72568	Sequence 72568, A
27	73	14.1	499	13	US-10-027-632-72567	Sequence 72567, A
28	73	14.1	499	14	US-10-027-632-72568	Sequence 72568, A
29	73	14.1	499	13	US-10-027-632-50724	Sequence 50724, A
30	73	14.1	499	14	US-10-027-632-50724	Sequence 50724, A
31	73	14.1	2828	13	US-10-024-298A-132	Sequence 132, App
32	72.5	14.0	2828	13	US-10-042-211A-132	Sequence 132, App
33	72	14.0	536165	11	US-09-939-964-1	Sequence 1, Appli
34	72	13.9	877	13	US-10-316-253-288	Sequence 288, Appl
35	71	13.7	889	13	US-10-316-253-286	Sequence 286, Appl
36	71	13.7	470	9	US-09-864-761-10114	Sequence 10114, A
37	70.5	13.6	2048	12	US-10-108-260A-2016	Sequence 2016, Ap
38	70.5	13.6	691	12	US-10-027-632-141792	Sequence 141792,
39	70	13.5	691	14	US-10-027-632-141792	Sequence 141792,
40	70	13.5	1416	13	US-09-887-586A-49	Sequence 19, Appl
41	69.5	13.4	1785	13	US-09-025-145A-19	Sequence 49, Appl
42	69.5	13.4	1785	9	US-09-903-012-49	Sequence 49, Appl
43	69.5	13.4	1785	11	US-09-900-797-49	Sequence 49, Appl
44	69.5	13.4	1785	14	US-10-041-007-19	Sequence 19, Appli
45	69	13.3	773	15	US-10-198-846-5618	Sequence 5618, Ap

RESULT 1
US-10-001-254-5
Sequence 5, Application US/10001254.
Publication No. US2003004702A1
GENERAL INFORMATION:
APPLICANT: Reed, John C.
APPLICANT: Godzik, Adam
APPLICANT: Pawlowski, Krzysztof
APPLICANT: Fiorentino, Loredana
APPLICANT: Lee, Sug Hyung
APPLICANT: Roth, Wilfred
APPLICANT: Stenner-Liewen, Frank
TITLE OF INVENTION: No. US2003004702A1e1 Death Domain Proteins
FILE REFERENCE: P-LJ 5037
CURRENT APPLICATION NUMBER: US/10/001,254
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: 60/301,889
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: 09/715,893
PRIOR FILING DATE: 2000-11-17
NUMBER OF SEQ ID NOS: 62
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 5
LENGTH: 294

ALIGNMENTS

```
;    TYPE: DNA
;    ORGANISM: Homo sapien
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (1)...(294)
US-10-001-254-5

Alignment Scores:
Pred. No.:
Score:                  1.93e-72
Percent Similarity:     517.00
Best Local Similarity:  100.00%
Query Match:            100.00%
DB:                     15

                        Length:        294
                        Matches:       98
                        Conservative:  0
                        Mismatches:    0
                        Indels:        0
                        Gaps:          0

US-10-001-254-6 (1-98) x US-10-001-254-5 (1-294)

Qy    1  ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ACATATGTCCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT  60

Qy   21  GlnGluGlyTrpLysLeuAlaValAlaIleIleLysLysProSerGlyAspAspArgTyr  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CAAGAGGAGGATGGAAGAAGTTAGCTGTAGCTATTATAAAAAAACCATCTGGTGATGATAGATAC  120

Qy   41  AsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrGlyLysSerProThrSer  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGTAAAAAGTCCCACTTCT  180

Qy   61  GlyLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  GAATTACTGTTTGACTGGGGCCACCACAAAATTGCACAGTTGGTGGATCTTTTG  240

Qy   81  IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro  98
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC  294
```

RESULT 2
US-10-001-254-15
; Sequence 15, Application US/1000254
; Publication No. US20030049702A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sug Hyung
; APPLICANT: Roth, Wilfried
; APPLICANT: Stenner-Liewen, Frank
; TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins
; FILE REFERENCE: P-LJ 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1383
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1380)

Alignment Scores:
Pred. No.:
Score: 1.8e-71
Percent Similarity: 517.00
Best Local Similarity: 100.00%

                        Length: 1383
                        Matches: 98
                        Conservative: 0
                        Mismatches: 0

```
Query Match:            100.00%        Indels:   0
DB:                     15             Gaps:     0

US-10-001-254-6 (1-98) x US-10-001-254-15 (1-1383)

Qy    1  ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25  ACATATGTCCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT  84

Qy   21  GlnGluGlyTrpLysLeuAlaValAlaIleIleLysLysProSerGlyAspAspArgTyr  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85  CAAGAGGAGGATGGAAGAAGTTAGCTGTAGCTATTATAAAAAAACCATCTGGTGATGATAGATAC  144

Qy   41  AsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrGlyLysSerProThrSer  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  145  AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGTAAAAAGTCCCACTTCT  204

Qy   61  GlyLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  205  GAATTACTGTTTGACTGGGGCCACCACAAAATTGCACAGTTGGTGGATCTTTTG  264

Qy   81  IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro  98
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  265  ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC  318
```

RESULT 3
US-09-966-451-3
; Sequence 3, Application US/09966451
; Publication No. US20030087856A1
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF IL-1 RECEPTOR-ASSOCIATED KINASE-4 EXPRESS
; FILE REFERENCE: RTS-0324
; CURRENT APPLICATION NUMBER: US/09/966,451
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 3
; LENGTH: 2817
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (50)...(1432)
US-09-966-451-3

Alignment Scores:
Pred. No.:
Score: 5.01e-71
Percent Similarity: 517.00
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 11

                        Length: 2817
                        Matches: 98
                        Conservative: 0
                        Mismatches: 0
                        Indels: 0
                        Gaps: 0

```
US-10-001-254-6 (1-98) x US-09-966-451-3 (1-2817)

Qy    1  ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74  ACATATGTCCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT  133

Qy   21  GlnGluGlyTrpLysLeuAlaValAlaIleIleLysLysProSerGlyAspAspArgTyr  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134  CAAGAGGAGGATGGAAGAAGTTAGCTGTAGCTATTATAAAAAAACCATCTGGTGATGATAGATAC  193

Qy   41  AsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrGlyLysSerProThrSer  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  194  AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGTAAAAAGTCCCACTTCT  253

Qy   61  GlyLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  254  GAATTACTGTTTGACTGGGGCCACCACAAAATTGCACAGTTGGTGGATCTTTTG  313

Qy   81  IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro  98
```

```
Db   314 ATCCAAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTGTTCCC 367
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 4

US-10-001-254-27

; Sequence 27, Application US/10001254
; Publication No. US20030049702A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sug Hyung
; APPLICANT: Roth, Wilfred
; APPLICANT: Stenner-Liewen, Frank
; TITLE OF INVENTION: No. US2003004702A1e1 Death Domain Proteins
; FILE REFERENCE: P-LJ 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 2817
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (50)...(1429)
US-10-001-254-27

Alignment Scores:
Pred. No.:
Score: 517.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB:

US-10-001-254-6 (1-98) x US-10-001-254-27 (1-2817)

Length: 2817
Matches: 98
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

```
Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgGlyLysLeuSerAspPheIleAspPro 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74 ACATATGTCCGCTGCCTCAATGTTGGACTTATTAGGAAGCTGTCAGATTTTATTGATCCT 133

Qy   21 GlnGluGlyTrpLysGlyLeuAlaValAlaIleLeuLysLysProSerGlyAspAspArgTyr 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134 CAAGAGGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 193

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  194 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 253

Qy   61 GlyLeuLeuPheAspTrpPheGlyTyrThrThrAsnGlyValGlyAspLeuValAspLeuLeu 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  254 GGAATTACTGTTTGACTGGGGCCACCACAAATTGCACAGTTGGGTGATCTTGGATCTTTG 313

Qy   81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  314 ATCCAAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTGTTCCC 367
```

RESULT 5

US-09-759-595-2

; Sequence 2, Application US/09759595
; Publication No. US20030059916A1
; GENERAL INFORMATION:
; APPLICANT: Wesche, Holger
; APPLICANT: Li, Shyun
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: IRAK-4: Compositions and Methods of Use
; FILE REFERENCE: 018781-0039110US
; CURRENT APPLICATION NUMBER: US/09/759,595
; CURRENT FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 60/176,395
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1383
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human IL-1 receptor-associated kinase 4 (IRAK-4)
; OTHER INFORMATION: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1383)
US-09-759-595-2

Alignment Scores:
Pred. No.:
Score: 513.00
Percent Similarity: 98.98%
Best Local Similarity: 98.98%
Query Match: 99.23%
DB:

US-10-001-254-6 (1-98) x US-09-759-595-2 (1-1383)

Length: 1383
Matches: 97
Conservative: 0
Mismatches: 1
Indels: 0
Gaps: 0

```
Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgGlyLysLeuSerAspPheIleAspPro 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACATATGTGCGCTGCCTCAATGTTGGACTTATTAGGAAGCTGTCAGATTTTATTGATCCT 84

Qy   21 GlnGluGlyTrpLysGlyLeuAlaValAlaIleLeuLysLysProSerGlyAspAspArgTyr 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAAGAGGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 144

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  145 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 204

Qy   61 GlyLeuLeuPheAspTrpPheGlyTyrThrThrAsnGlyValGlyAspLeuValAspLeuLeu 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  205 GGAATTACTGTTTGACTGGGGCCACCACAAATTGCACAGTTGGGTGATCTTGGATCTTTG 264

Qy   81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  265 ATCCAAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTGTTCCC 318
```

RESULT 6

US-09-759-595-4

; Sequence 4, Application US/09759595
; Publication No. US20030059916A1
; GENERAL INFORMATION:
; APPLICANT: Wesche, Holger
; APPLICANT: Li, Shyun
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: IRAK-4: Compositions and Methods of Use
; FILE REFERENCE: 018781-0039110US
; CURRENT APPLICATION NUMBER: US/09/759,595
; CURRENT FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 60/176,395
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: murine IL-1 receptor-associated kinase 4 (IRAK-4)
; OTHER INFORMATION: cDNA

```
;      FEATURE:
;         NAME/KEY: CDS
;         LOCATION: (163)..(1542)
;         OTHER INFORMATION:
US-09-759-595-4
;         OTHER INFORMATION: murine IRAK-4

Alignment Scores:
Pred. No.:
Score:                      6.76e-65
                            476.00
Percent Similarity:         95.92%
Best Local Similarity:      90.82%
Query Match:                92.07%
DB:                         11

Length:                     1542
Matches:                    89
Conservative:               5
Mismatches:                 4
Indels:                     0
Gaps:                       0

US-10-001-254-6 (1-98) x US-09-759-595-4 (1-1542)

Qy   1  ThrTyrValArgCysLeuAsnValGlyTyrLeuIleArgGlyLeuLeuSerAspPheIleAspPro  20
Db 187  ACATACGTGCGCAACCTTAATGTGGGAGATCCTTAAGGGAAGCTGTCGGAGATTTATTGATCCT  246

Qy  21  GlnGluGlyTrpLysLysLeuAlaValAlaIleLeuLysLysProSerGlyAspAlaProArgTyr  40
Db 247  CAAGAAGGATGGAAGAAATTAGCAGTAGCTATCAAAAAAGCCTTCCGGCGACGACAGATAC  306

Qy  41  AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer  60
Db 307  AATCAGTTCCATATAAGGAGAATTCGAAGCCTTACTTCAGACCGGGAAGAGCCCCACTTGT  366

Qy  61  GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
Db 367  GAACTCGTGTTTGACTGGGGCACCACGAACTGCACAGTTGGGGATCTTGTTGGATCTACTG  426

Qy  81  IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuLeuLeuProAspAlaValPro  98
Db 427  GTCCAGAATTGAGCTGTTTGCCCCGGCCACTCTCCTGCTTCTCCTGCCGGATGCCGTTCCC  480


RESULT 7
US-09-833-790-149
; Sequence 149, Application US/0983790
; Patent No. US2002068288A1
; GENERAL INFORMATION:
;   APPLICANT: Lodes, Michael J.
;   APPLICANT: Wang, Tongtong
;   APPLICANT: Secrist, Heather
;   APPLICANT: Mohamath, Raodoh
;   APPLICANT: Indiras, Carol Y.
;   APPLICANT: Fan, Liqun
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;   TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
;   FILE REFERENCE: 210121.512
;   CURRENT APPLICATION NUMBER: US/09/833,790
;   CURRENT FILING DATE: 2001-04-11
;   NUMBER OF SEQ ID NOS: 440
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149
;   LENGTH: 501
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: (1)..(501)
;     OTHER INFORMATION: n = A,T,C or G
US-09-833-790-149

Alignment Scores:
Pred. No.:
Score:                      1.33e-55
                            413.00
Percent Similarity:         94.00%
Best Local Similarity:      93.00%
Query Match:                79.88%
DB:                         9

Length:                     501
Matches:                    93
Conservative:               1
Mismatches:                 4
Indels:                     4
Gaps:                       0

US-10-001-254-6 (1-98) x US-09-833-790-149 (1-501)


RESULT 8
US-09-966-451-10
; Sequence 10, Application US/09966451
; Publication No. US2003087856A1
; GENERAL INFORMATION:
;   APPLICANT: C. Frank Bennett
;   APPLICANT: Susan M. Freier
;   TITLE OF INVENTION: ANTISENSE MODULATION OF IL-1 RECEPTOR-ASSOCIATED KINASE-4 EXPRESS
;   FILE REFERENCE: RTS-0324
;   CURRENT APPLICATION NUMBER: US/09/966,451
;   CURRENT FILING DATE: 2001-09-28
;   NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 10
;   LENGTH: 31000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-966-451-10

Alignment Scores:
Pred. No.:
Score:                      4.2e-28
                            256.00
Percent Similarity:         98.00%
Best Local Similarity:      100.00%
Query Match:                49.52%
DB:                         11

Length:                     31000
Matches:                    49
Conservative:               0
Mismatches:                 0
Indels:                     0
Gaps:                       0

US-10-001-254-6 (1-98) x US-09-966-451-10 (1-31000)

Qy   1  ThrTyrValArgCysLeuAsnValGlyTyrLeuIleArgGlyLeuLeuSerAspPheIleAspPro  20
Db  45  ACATATGTGCGCTGCCTCAATGTTGGACTAATTAAGGGAACTGTCAGAGATTTTAATTGATCCT  104

Qy  21  GlnGluGlyTrpLysLysLeuAlaValAlaIleLeuLysLysProSerGlyAspAlaProArgTyr  40
Db 105  CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCAATCTGGTTGACATAGAATAC  164

Qy  41  AsnGln-PheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer  60
Db 165  AATCAAGTTTCACATAAGGAGAGATTTGAAGCATT-CTTCAAACTGGGAAAAAGTCCCACTTC  223

Qy  61  GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
Db 224  TTGAATACTGTTTGACTGGGGCACCACCAAATTGGACAGTTGGTGATCTTGTGGATCTTT  283

Qy  81  IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuLeuLeuPro  94
Db 284  TGATCCAAAATGAATT-TTTGCTCCTGCAAGCTCTTTGCTCCCAGATCGTCGTTCC  338


RESULT 9
US-10-001-254-25
; Sequence 25, Application US/10001254
; Publication No. US2003049702A1
; GENERAL INFORMATION:
;   APPLICANT: Reed, John C.
;   APPLICANT: Godzik, Adam
;   APPLICANT: Pawlowski, Krzysztof
;   APPLICANT: Fiorentino, Loredana
;   APPLICANT: Lee, Sug Hyung
;   APPLICANT: Roth, Wilfred
```

Alignment Scores:
Pred. No.:
Score: 4.2e-28 / 256.00
Percent Similarity: 100.00%
Best Local Similarity: 98.00%
Query Match: 49.52%

```
US-10-001-254-6 (1-98) x US-09-966-451-10 (1-31000)

Qy    45  IleLeuArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPhe  64
Db 12963  TTAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTT  13022

Qy    65  AspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsnGlu  84
Db 13023  GACTGGGGCACCACAAATTGCACAGTTGGTGATCTTTTGTGGATCCAAAATGAA  13082

Qy    85  PhePheAlaProAlaSerLeuLeuLeuLeuLeuPro  94
Db 13083  TTTTTTGCTCCTGCAAGCTCTTTTGCTCCCA  13112
```

```
;   APPLICANT:  Stenner-Liewen, Frank
;   TITLE OF INVENTION: No. US20030049702A1el Death Domain Proteins
;   FILE REFERENCE: P-LJ 5037
;   CURRENT APPLICATION NUMBER: US/10/001,254
;   CURRENT FILING DATE: 2001-11-15
;   PRIOR APPLICATION NUMBER: 60/301,889
;   PRIOR FILING DATE: 2001-06-29
;   PRIOR APPLICATION NUMBER: 09/715,893
;   PRIOR FILING DATE: 2000-11-17
;   NUMBER OF SEQ ID NOS: 62
;   SOFTWARE: FastSEQ for Windows Version 4.0
    SEQ ID NO 25
;   LENGTH: 211
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(177)
US-10-001-254-25

Alignment Scores:
Pred. No.:
Score:                    8.48e-30      Length:          211
Percent Similarity:       247.00        Matches:          46
Best Local Similarity:    100.00%       Conservative:      0
Query Match:              100.00%       Mismatches:        0
                          47.78%        Indels:            0
                          15            Gaps:              0

US-10-001-254-6 (1-98) x US-10-001-254-25 (1-98)

Qy        1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       25 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAGCTGTCAGATTTTATTGATCCT 84

Qy       21 GlnGlyGlyTrpLysLeuValAlaValAlaIleLeuLysLysProSerGlyAspAspArgTyr 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       85 CAAGGAGGATGGAAGAAAGTTAGCCTGTAGCTATTAAAAAAACCATCTCGGTGATGATAGATAC 144

Qy       41 AsnGlnPheHisIleArg 46
            ||||||||||||||||
Db      145 AATCAGTTTCACATAGA 162
```

```
RESULT 10
US-09-863-549-1
;   Sequence 1, Application US/09865549
;   Patent No. US20020049300A1
;   GENERAL INFORMATION:
;   APPLICANT: Cao, Zhaodan
;   TITLE OF INVENTION: IRAK3 Polypeptides, Polynucleotides and Methods
;   FILE REFERENCE: T98-019
;   CURRENT APPLICATION NUMBER: US/09/863,549
;   CURRENT FILING DATE: 2001-05-22
;   PRIOR APPLICATION NUMBER: 09/135,232
;   PRIOR FILING DATE: 1998-08-17
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: Patentin Ver. 2.0
    SEQ ID NO 1
;   LENGTH: 2288
;   TYPE: DNA
;   ORGANISM: human
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (64)..(1851)
US-09-863-549-1

Alignment Scores:
Pred. No.:
Score:                    0.000202      Length:          2288
Percent Similarity:       96.50         Matches:           23
Best Local Similarity:    47.95%        Conservative:      12
Query Match:              31.51%        Mismatches:        31
                          18.67%        Indels:             7
                          9             Gaps:               1
```

```
US-10-001-254-6.rnpb

US-10-001-254-6 (1-98) x US-09-863-549-1 (1-2288)

Qy        8 ValGlyLeuIleArgLysLeuSerAspPheIleAspProGlnGlyGlyTrpLysLeu 27
            :::  ||| :::  ||| ::  |||     :::  |||||| ::
Db      136 CTCGGAGAGCTCGTGCGCGTTCTTCGGACACAGCTGCGACGCGGCGCGGCGGCGCGGCTG 195

Qy       28 AlaValAlaIleLeuLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArg 47
            ::                    -------           ::|||
Db      196 GCAGAGAGACTTTCAAGC--------------------AGCTTGGCTGGATGTTCGTCAT 234

Qy       48 PheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGlyLeuLeuPheAspTrpGly 67
            |||          |:::  |:::  |||||||:::                 :::  |||
Db      235 ATTGAAAAGTTATGTAGACCAAGGTAAAAAGTGGAACAAGAGAATTACTTTGGTCCTGGGCA 294

Qy       68 ThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
            |||            :::::  :::||| :::  |||:::
Db      295 CAGAAAAACAACGACCATCGGTGACCTTTACAGGTCCTC 333
```

```
RESULT 11
US-10-098-841-3
;   Sequence 3, Application US/10098841
;   Publication No. US20020197679A1
;   GENERAL INFORMATION:
;   APPLICANT: Tang, Y. Tom
;   APPLICANT: Liu, Chenghua
;   APPLICANT: Asundi, Vinod
;   APPLICANT: Wang, Dunrui
;   APPLICANT: Wang, Zhiwei
;   APPLICANT: Wehrman, Tom
;   APPLICANT: Zhang, Jie
;   APPLICANT: Qian, Xiaohong B.
;   APPLICANT: Drmanac, Radoje T.
;   APPLICANT: Xu, Chongjun
;   APPLICANT: Zhou, Ping
;   APPLICANT: Ma, Yunqing
;   APPLICANT: Wang, Jian-Rui
;   APPLICANT: Zhao, Qing A.
;   APPLICANT: Ren, Feiyan
;   APPLICANT: Chen, Rui-hong
;   TITLE OF INVENTION: 784CIP2
;   TITLE OF INVENTION: Polypeptides
;   FILE REFERENCE: US/10/098,841
;   CURRENT APPLICATION NUMBER: 2002-03-13
;   CURRENT FILING DATE: 09/598,042
;   PRIOR APPLICATION NUMBER: 2000-06-20
;   PRIOR FILING DATE: 09/552,317
;   PRIOR APPLICATION NUMBER: 2000-04-25
;   PRIOR FILING DATE: 09/488,725
;   PRIOR APPLICATION NUMBER: 2000-01-21
;   PRIOR FILING DATE: 331
;   NUMBER OF SEQ ID NOS: pt_FL_genes Version 1.0
;   SOFTWARE: US20020197679A1el Nucleic Acids and
    SEQ ID NO 3
;   LENGTH: 2293
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (7)..(1806)
US-10-098-841-3

Alignment Scores:
Pred. No.:
Score:                    0.000203      Length:          2293
Percent Similarity:       96.50         Matches:           23
Best Local Similarity:    47.95%        Conservative:      12
Query Match:              31.51%        Mismatches:        31
                          18.67%        Indels:             7
                          14            Gaps:               1

US-10-001-254-6 (1-98) x US-10-098-841-3 (1-2293)

Qy        8 ValGlyLeuIleArgLysLeuSerAspPheIleAspProGlnGlyGlyTrpLysLeu 27
            ::||||      :::  |||           |||  |||||| :::       |||
```

```
Db    88  CTCGGAGAGCTCTGCGCGTGTTCTGGACAGCTGCCAGCGCACGGCGCTGGGCGGCGGGCCTG  147
Qy    28  AlaValAlaIleLeuLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArg  47
                                                                    :::
Db   148  GCAGAGAGACTTTCAAGC---                                          186
Qy    48  PheGluValAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGly  67
          |||                              ||||||::
Db   187  ATTGAAAAGTATGTGACCAAGGTAAAAGTGGAACAAGAGAATTACTTTGGTCCTGGGGCA    246
Qy    68  ThrThrAsnCysThrValGlyAspLeuValAspLeuLeu                        80
          |||::||||||   |||    :::|||
Db   247  CAGAAAAACAAGACCATCGGTGACCTTTACAGGTCCTC                        285
```

RESULT 12
US-10-340-545-1
; Sequence 1, Application US/10340545
; Publication No. US20030157539A1
; GENERAL INFORMATION:
; APPLICANT: Flavell et al.
; TITLE OF INVENTION: IRAK-M IS A NEGATIVE REGULATOR OF TOLL-LIKE RECEPTOR SIGNALING
; FILE REFERENCE: YU-P01-011
; CURRENT APPLICATION NUMBER: US/10/340,545
; CURRENT FILING DATE: 2003-01-09
; PRIOR APPLICATION NUMBER: 60/348176
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1888
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-340-545-1

Alignment Scores:
Pred. No.:
Score: 0.000265
                       95.00
Percent Similarity: 48.65%
Best Local Similarity: 32.43%
Query Match: 18.38%
DB: 13

US-10-001-254-6 (1-98) x US-10-340-545-1 (1-1888)

                Length: 1888
                Matches: 24
                Conservative: 12
                Mismatches: 28
                Indels: 10
                Gaps: 2

Qy   10 LeuIleLeuArgLysLeuSerAspPheIleAspProGlnGlu---------GlyTrpLysLys 26
          ||:: || ::||||
Db   128 CTGCTGGGGAGGAGCTTTGGCGGGGATCCTGGACAGTGGGGATGCCCGCCTCGGCTGGTGGGGC 187

Qy   27 LeuAlaValAlaIleLeuLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 46
          ||| ||| ::
Db   188 CTGGCGGGGACGGACTTTCAAAC----------AGCTGGCTGGAGATGTTCGT   226

Qy   47 ArgPheGluValAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrp 66
          ||| :::||||||| |||
Db   227 CATATTGAAAAGTACCTAACCAAGGTAAAAGTGGAACAAGAGAATTGCTCTGGTCCTGG   286

Qy   67 GlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu   80
          ||:::|||||| ||| :::|||
Db   287 GCACAGAAAAACAAAACGATCGGTCGGCCACCTTTAGAGGTTCTC   328
```

; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1432
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2, 7, 217
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-1432

Alignment Scores:
Pred. No.:
Score:                 0.0105
                       79.50
Percent Similarity:    45.45%
Best Local Similarity: 27.27%
Query Match:           15.38%
DB:                    15

US-10-001-254-6 (1-98) x US-10-198-846-1432 (1-476)

                Length:        476
                Matches:       24
                Conservative:  16
                Mismatches:    37
                Indels:        11
                Gaps:          3

Qy     2  TyrValArgCysLeuAsnValGlyLeu-------------------------IleArgLys  13
          ::||                                              ::::::||
Db   154  TACTCGAGGCGCCTTGATCTCTGGCCTCTTGCCATCCCTCCCATTATTCTGAAGAA      213

Qy    14  LeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLeuLysLys  33
          :::||                          |||        |||
Db   214  GTACCCCATTTTTTGCAGATTCCGGAGGGAGATGGGGA---------GCCACGGCGCCGCTCTAGA  267

Qy    34  ProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluValAlaLeuLeuGln  53
          |||:::          ::||||||            |||
Db   268  ACTAGTGGATCCCCGGGCTGCACGAAATTCGAATATCAAGCTTATCGAGATACCGTCGACTC  327

Qy    54  ThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThrVal  73
          |||           :::||||||              |||
Db   328  GAGGGG---GGGCCCGGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGTT  384

Qy    74  GlyAspLeuValAspLeuLeuIle  81
          ||:::||||     :::|||
Db   385  GGGGTAATCATGGTCATAGCTGTT  408
```

RESULT 13
US-10-198-846-1432
; Sequence 1432, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049

RESULT 14
US-10-366-288-19
; Sequence 19, Application US/10366288
; Publication No. US20030216288A1
; GENERAL INFORMATION:
; APPLICANT: Powell, Douglas
; APPLICANT: Welch, Nadine S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING
; TITLE OF INVENTION: AIDS AND HIV-RELATED DISORDERS USING
; TITLE OF INVENTION: 34021, 1720, 1683, 1552, 1682, 1675, 1481, 1553,
; TITLE OF INVENTION: 10002, 1611, 1371, 14324, 126, 270, 312, 167, 18926,
; TITLE OF INVENTION: 6747, 1793, 1784 OR 2045 MOLECULES
; FILE REFERENCE: MPI02-025P1RNONNIM
; CURRENT APPLICATION NUMBER: US/10/366,288
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: 60/357,391
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/380,249
; PRIOR FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: 60/391,306
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: 60/406,297
; PRIOR FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: 60/412,007
; PRIOR FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: 60/417,508
; PRIOR FILING DATE: 2002-10-10

; PRIOR APPLICATION NUMBER: 60/432,318
; PRIOR FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 19
; LENGTH: 1782
; TYPE: DNA
; ORGANISM: Homo Sapien
;
US-10-366-288-19

Alignment Scores:
Pred. No.:
Score: 0.0844
Percent Similarity: 79.00
Best Local Similarity: 49.28%
Query Match: 31.88%
DB: 15.28%
                                13

US-10-001-254-6 (1-98) x US-10-366-288-19 (1-1782)

Length: 1782
Matches: 22
Conservative: 12
Mismatches: 19
Indels: 16
Gaps: 2

Qy   45 IleArgArgPheGluAlaLeuLeuGlnThr---GlyLysSerProThrSerGluLeuLeu 63
            ::|:||| |:   ::: || ||| ||||||||||||
Db   133 CTGCGGAAGATCAAGTCCATGGAGCGGGGTGCAGGGGTGTGAGGAGCATCACGCGGGGAGCTGCTG 192

Qy   64 PheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsn 83
            |||||| | |||||||||||| ||| | |:|||||||| :::
Db   193 TGGTGGTGGGGCGGGGCCATGCGGCGCCACCGTTCCAGCAACTTGTGGACCTTCCTGTGCCGCGTG 252

Qy   84 GluPhePheAlaProAlaSerLeuLeuLeu------------------------ 93
            ||| | | ||| | :|:::||
Db   253 GAGCTGTACCGGCGCTCCCCAGATCATCCTGAACTGGAAAACCGGCTCCTGAAATCAGGTGT 312

Qy   94 ---------------------------ProAspAlaVal 97
                                        |||||||||:::||
Db   313 CCCATTCCAGCCTTCCCTGACTCTGTG 339

RESULT 15
US-09-773-753-1
; Sequence 1; Application US/0973753
; Patent No. US20020091777A1
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
;   FENG, PING
;   MUZIO, MARTA
;   DIXIT, VISHVA M.
; TITLE OF INVENTION: HUMAN IRAK-2
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/773,753
; FILING DATE: 02-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/980,060
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1806 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..1803
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-773-753-1

Alignment Scores:
Pred. No.:
Score: 0.086
Percent Similarity: 79.00
Best Local Similarity: 49.28%
Query Match: 31.88%
DB: 15.28%
                                9

US-10-001-254-6 (1-98) x US-09-773-753-1 (1-1806)

Length: 1806
Matches: 22
Conservative: 12
Mismatches: 19
Indels: 16
Gaps: 2

Qy   45 IleArgArgPheGluAlaLeuLeuGlnThr---GlyLysSerProThrSerGluLeuLeu 63
            ::|:||| |:   ::: || ||| ||||||||||||
Db   157 CTCCGGAAGATCAAGTCCATGGAGCGGGGTGCAGGGGTGTGAGGAGCATCACGCGGGGAGCTGCTG 216

Qy   64 PheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuIleGlnAsn 83
            |||||| | |||||||||||| ||| | |:|||||||| :::
Db   217 TGGTGGTGGGGCATGCGGCGCCACCGTTCCAGCAACTTGTGGACCTTCCTGTGCCGCGTG 276

Qy   84 GluPhePheAlaProAlaSerLeuLeuLeu------------------------ 93
            ||| | | ||| | :|:::||
Db   277 GAGCTGTACCGGCGCTCCCCAGATCATCCTGAACTGGAAAACCGGCTCCTGAAATCAGGTGT 336

Qy   94 ---------------------------ProAspAlaVal 97
                                        |||||||||:::||
Db   337 CCCATTCCAGCCTTCCCTGACTCTGTG 363

Search completed: January 18, 2004, 09:05:28
Job time : 153.603 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 18, 2004, 01:11:43 ; Search time 956.969 Seconds
(without alignments)
2488.940 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

```
Title:           US-10-001-254-6
Perfect score:   517
Sequence:        1 TVVRCLNVGLIRKLSDFIDP.............LLIQNEFPAPASLLPDAVP 98

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        22781392 seqs, 1215223B056 residues

Total number of hits satisfying chosen parameters:  45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152424_19723/app_query.fasta_1.1109
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10001254_@CGN_1_14382_@runat_16012004_152424_19723 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       EST:*
                 1:  em_estba:*
                 2:  em_esthum:*
                 3:  em_estin:*
                 4:  em_estmu:*
                 5:  em_estov:*
                 6:  em_estpl:*
                 7:  em_estro:*
                 8:  em_htc:*
                 9:  em_est1:*
                 10: gb_est2:*
                 11: gb_htc:*
                 12: gb_est3:*
                 13: gb_est4:*
                 14: gb_est5:*
                 15: em_estfun:*
                 16: em_estom:*
                 17: em_estnum:*
                 18: em_gss_hum:*
                 19: em_gss_inv:*
                 20: em_gss_pln:*
                 21: em_gss_vrt:*
                 22: em_gss_fun:*
                 23: em_gss_mam:*
                 24: em_gss_mus:*
                 25: em_gss_pro:*
                 26: em_gss_rod:*
                 27: em_gss_phg:*
                 28: gb_gss1:*
                 29: gb_gss2:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	517	100.0	859	10	BG164491	BG164491 602342026
2	509	98.5	811	10	BG616438	BG616438 602642772
3	495	95.7	402	10	BB482619	BB482619 168463 BA
4	485	93.8	719	10	BF696981	BF696981 602130160
5	476	92.1	503	14	CA538859	CA538859 C0272B01-
6	476	92.1	598	13	BQ552228	BQ552228 H4014C09-
7	476	92.1	610	10	BB660378	BB660378 BB660378
8	476	92.1	637	10	BB613447	BB613447 BB613447
9	476	92.1	638	14	BY721552	BY721552 BY721552
10	476	92.1	663	14	BY726858	BY726858 BY726858
11	476	92.1	676	10	BB613167	BB613167 BB613167
12	476	92.1	1161	11	AK020397	AK020397 Mus muscu
13	476	92.1	2481	11	AK028837	AK028837 Mus muscu
14	476	92.1	2810	11	AK029028	AK029028 Mus muscu
15	472	91.3	575	4	BX522921	BX522921 RZPD Mus
16	430.5	83.3	524	10	AW106160	AW106160 um23h11.y
17	415	80.3	541	10	BG691069	BG691069 340084 BA
18	414.5	80.2	507	10	BB866698	BB866698 BB866698
19	393	76.0	453	10	BB860349	BB860349 BB860349
20	393	76.0	852	13	BU209111	BU209111 603950834
21	389	75.2	600	9	AJ453616	AJ453616 AJ453616
22	389	75.2	670	9	AJ447581	AJ447581 AJ447581
23	369	71.4	265	9	AW436511	AW436511 76774 MAR
24	349	67.5	313	12	BM151935	BM151935 TCBAP1E11
25	343	66.3	629	9	AL647125	AL647125 AL647125
26	323.5	62.6	664	12	BJ035962	BJ035962 BJ035962
27	296	57.3	632	14	CA365604	CA365604 640775 NC
28	289	55.9	743	14	CA474136	CA474136 AGENCOURT
29	287	55.5	555	9	AW423082	AW423082 f168b07.Y
30	270	52.2	520	10	BE132064	BE132064 db41h09.y
31	264	51.1	380	12	BI883790	BI883790 fn07c09.y
32	258	49.9	505	10	BE482323	BE482323 HS64 BA
33	247	47.8	284	9	AA114228	AA114228 zn75g05.r
34	247	47.8	540	9	AL699213	AL699213 DKP2p686K
35	244	47.2	576	10	BF238344	BF238344 601904613
36	237	45.8	781	10	BF723681	BF723681 EST00673
37	236	45.6	795	13	BQ293310	BQ293310 EST01080
38	226	43.7	270	10	BB845841	BB845841 232716 BA
39	204	39.5	490	10	AJ443799	AJ443799 AJ443799
40	193	37.3	858	10	BF687921	BF687921 602069996
41	132	25.5	318	14	CB940191	CB940191 IpCGJx14
42	124	24.0	595	28	AQ601806	AQ601806 HS_2108_B
43	121.5	23.5	476	9	AV976730	AV976730 AV976730
44	121.5	23.5	509	9	AV995166	AV995166 AV995166
45	121.5	23.5	567	12	BP015417	BP015417 BP015417

ALIGNMENTS

```
RESULT 1
BG164491
LOCUS       BG164491                 859 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602342026F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4452055 5',
            mRNA sequence.
ACCESSION   BG164491
VERSION     BG164491.1  GI:12671194
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 859)
```

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
　　　　　National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
　　　　　Email: cgapbs-r@mail.nih.gov
　　　　　Tissue Procurement: ATCC
　　　　　cDNA Library Preparation: Life Technologies, Inc.
　　　　　cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
　　　　　DNA Sequencing by: Incyte Genomics, Inc.
　　　　　Clone distribution: MGC clone distribution information can be
　　　　　found through the I.M.A.G.E. Consortium/LLNL at:
　　　　　http://image.llnl.gov
　　　　　Plate: LLAM10240 row: d column: 08
　　　　　High quality sequence stop: 634.
FEATURES　　　　　Location/Qualifiers
　　　　source　　　　1..859
　　　　　　　　/organism="Homo sapiens"
　　　　　　　　/mol_type="mRNA"
　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　/clone="IMAGE:4452055"
　　　　　　　　/tissue_type="hypernephroma, cell line"
　　　　　　　　/lab_host="DH10B (phage-resistant)"
　　　　　　　　/clone_lib="NIH_MGC_89"
　　　　　　　　/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
　　　　　　　　Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
　　　　　　　　Average insert size 1.3 kb. Library enriched for
　　　　　　　　full-length clones and constructed by Life Technologies.
　　　　　　　　Note: this is a NIH_MGC Library."
BASE COUNT　　266 a　　168 c　　190 g　　235 t
ORIGIN

US-10-001-254-6 (1-98) x BG164491 (1-859)

Alignment Scores:
Pred. No.:
Score:　　　　　1.21e-61　　　Length:　　　859
Percent Similarity:　517.00　　Matches:　　859
Best Local Similarity:　100.00%　Conservative:　98
Query Match:　100.00%　　Mismatches:　0
DB:　　　　　　100.00%　　Indels:　　0
　　　　　　　　　　　　　　　Gaps:　　　0
　　　　　　　　　　　　　　10

```
Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   59 ACATATGTCCGCTGCCTCAATGTTGGACTAATTAGGAAGCTTGTCAGATTTTATTGATCCT 118

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 178

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  179 AATCAGTTTCACATAAAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 238

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  239 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 298

Qy   81 IleGlnAsnGlyPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  299 ATCCAAAATGGAATTTTTCGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 352
```

RESULT 2
BG616438
LOCUS　　　BG616438　　　811 bp　　　mRNA　　　linear　　　EST 18-APR-2001
DEFINITION 602642772F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4773760 5',
　　　　　　mRNA sequence.
ACCESSION　BG616438
VERSION　　BG616438.1　　GI:13667809
KEYWORDS　EST.
SOURCE　　Homo sapiens (human)
　ORGANISM　Homo sapiens
　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE　　1 (bases 1 to 811)
　AUTHORS　NIH-MGC http://mgc.nci.nih.gov/.
　　　　　National Institutes of Health, Mammalian Gene Collection (MGC)
　TITLE　Unpublished
　JOURNAL　Unpublished
COMMENT　Contact: Robert Strausberg, Ph.D.
　　　　　Email: cgapbs-r@mail.nih.gov
　　　　　Tissue Procurement: ATCC
　　　　　cDNA Library Preparation: CLONTECH Laboratories, Inc.
　　　　　cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
　　　　　DNA Sequencing by: Incyte Genomics, Inc.
　　　　　Clone distribution: MGC clone distribution information can be
　　　　　found through the I.M.A.G.E. Consortium/LLNL at:
　　　　　http://image.llnl.gov
　　　　　Plate: LLCM1645 row: p column: 17
　　　　　High quality sequence start: 3
　　　　　High quality sequence stop: 613.
FEATURES　　　　Location/Qualifiers
　　　　source　　　1..811
　　　　　　　　/organism="Homo sapiens"
　　　　　　　　/mol_type="mRNA"
　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　/clone="IMAGE:4773760"
　　　　　　　　/tissue_type="embryonal carcinoma"
　　　　　　　　/lab_host="DH10B (T1 phage-resistant)"
　　　　　　　　/clone_lib="NIH_MGC_61"
　　　　　　　　/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
　　　　　　　　Sfi (ggccgcctcggcc); Site_2: Sfi (ggccattaggcc);
　　　　　　　　Double-stranded cDNA was prepared from cell line RNA. 5'
　　　　　　　　and 3' adaptors were used in cloning as follows: 5'
　　　　　　　　adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
　　　　　　　　sequence: 5'-ATTCTAGAGGCGGCCGGGGGCCCCGACATG-GT-3'
　　　　　　　　(where B = A, C; or G and N = A, C, G, or T). Average
　　　　　　　　insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
　　　　　　　　contained inserts by PCR. This library was enriched for
　　　　　　　　full-length clones and was constructed by Clontech
　　　　　　　　Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
　　　　　　　　Library."
BASE COUNT　　246 a　　196 c　　164 g　　205 t
ORIGIN

US-10-001-254-6 (1-98) x BG616438 (1-811)

Alignment Scores:
Pred. No.:
Score:　　　　　1.47e-60　　　Length:　　　811
Percent Similarity:　509.00　　Matches:　　97
Best Local Similarity:　98.98%　Conservative:　0
Query Match:　98.98%　　Mismatches:　1
DB:　　　　　　98.45%　　Indels:　　0
　　　　　　　　　　　　　　　Gaps:　　　0
　　　　　　　　　　　　　　10

```
Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   77 ACATATGTCCGCTGCCTCAATGTTGGACTAATTAGGAAGCTTGTCAGATTTTATTGATCCT 136

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 196

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  197 AATCAGTTTCACATAAAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 256

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  257 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 316

Qy   81 IleGlnAsnGlyPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  317 ATCCAAAATGGAATTTTTCTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 370
```

RESULT 3
BE482619
LOCUS　　　BE482619　　　402 bp　　　mRNA　　　linear　　　EST 27-MAR-2003

Left column

```
DEFINITION   168463 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION    BE482619
VERSION      BE482619.1  GI:9602152
KEYWORDS     EST.
SOURCE       Bos taurus (cow)
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
             Bovidae; Bovinae; Bos.
REFERENCE    1  (bases 1 to 402)
  AUTHORS    Sonstegard,T., Capuco,A.V., White,J., Van Tassell,C.P., Connor,E.E.
             , Cho,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D. and
             Quackenbush,J.
  TITLE      Analysis of bovine mammary gland EST and functional annotation of
             the Bos taurus gene index
  JOURNAL    Mamm. Genome 13 (7), 373-379 (2002)
  MEDLINE    22135956
   PUBMED    12140684
COMMENT      Contact: Sonstegard TS
             USDA, ARS, Beltsville Agricultural Research Center
             Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
             Tel: 301 504 8416
             Fax: 301 504 8414
             Email: tads@lpsi.barc.usda.gov
             Single pass sequencing. Bases called and alt_trimmed with phred
             v0.980904.e. Vector identified by cross_match with the -minscore 18
             and -minmatch 12 options.
             PCR Primers
             FORWARD: AGGAAACAGCTATGACCAT
             BACKWARD: GTTTTCCCAGTCACGACG
             Plate: 11 row: F column: 7
             Seq primer: ATTTAGGTGACACTATAG.
FEATURES     Location/Qualifiers
     source  1..402
             /organism="Bos taurus"
             /mol_type="mRNA"
             /db_xref="taxon:9913"
             /tissue_type="pooled"
             /lab_host="DH10B"
             /clone_lib="BARC 5BOV"
             /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
             Library made from pooled mRNA isolated from mammary
             tissues at eight physiological, developmental, and disease
             states."
BASE COUNT   114 a   95 c   88 g   105 t
ORIGIN
```

US-10-001-254-6 (1-98) x BE482619 (1-402)

```
Alignment Scores:
Pred. No.:               1
Score:             5.03e-59     495.00     Length:        402
Percent Similarity:  97.96%                Matches:        92
Best Local Similarity: 93.88%              Conservative:    4
Query Match:         95.74%                Mismatches:      2
                                10         Indels:          0
                                           Gaps:            0
```

```
Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
         |||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db  88 ACATACGTGCGCTGCCTCAGTCCTTGGACTAATTAGGAAAGCTGTCCAGATTTTATTGATCCT 147

Qy  21 GlnGlyLeuTrpLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 148 CAAGGAGGATGGAAGAAGTTAGCAGTGCGTATTAAAAAACCATCTGGTGATGATAGATAC 207

Qy  41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 208 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTGCAAAATTGGAAAAAGCCCACGTGT 267

Qy  61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
         |||||||||||||||||||||||||||||||||||||||||||:|||::|||
Db 268 GAGTTACTGTTTGACTGGGGCACCACAAATTGCACAGTGGTGGATCTTGTTGGATAATTTG 327
```

Right column

```
Qy  81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuProAspAlaValPro 98
         :::|||||||||||||||||||||||||||||||||||||||||||||||||
Db 328 GTCCAAAATGAGTTTTTTGCCCCTGCAAGTCTTTTGCTACCAGATGCTGTACCC 381
```

```
RESULT 4
BF696981
  Db   BF696981   719 bp   mRNA   linear   EST 22-DEC-2000
LOCUS        BF696981   719 bp   mRNA   linear   EST 22-DEC-2000
DEFINITION   602130160F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4287014 5',
             mRNA sequence.
ACCESSION    BF696981
VERSION      BF696981.1  GI:11982389
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 719)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCMI124 row: g column: 15
             High quality sequence stop: 632.
FEATURES     Location/Qualifiers
     source  1..719
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:4287014"
             /tissue_type="primitive neuroectoderm"
             /lab_host="DH10B (T1 phage-resistant)"
             /clone_lib="NIH_MGC_56"
             /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
             SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
             Double-stranded cDNA was prepared from cell line RNA. 5'
             and 3' adaptors were used in cloning as follows: 5'
             adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
             sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACGCGTG-dT(30)BN-3'
             (where B = A, C, G or T, and N = A, C, G, or T). Average
             insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
             contained inserts by PCR. This library was enriched for
             full-length clones and was constructed by Clontech
             Laboratories (Palo Alto, CA)."
BASE COUNT   216 a   145 c   163 g   195 t
ORIGIN
```

US-10-001-254-6 (1-98) x BF696981 (1-719)

```
Alignment Scores:
Pred. No.:               1
Score:             2.87e-57     485.00     Length:        719
Percent Similarity:  97.00%                Matches:        97
Best Local Similarity: 97.00%              Conservative:    0
Query Match:         93.81%                Mismatches:      1
                                10         Indels:          2
                                           Gaps:            0
```

```
Qy   1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAspPro 20
         |||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db  92 ACATACGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 151

Qy  21 GlnGlyLeuTrpLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152 CAAGGAGGATGGAAGAAGTTAGCTGTCGTATTAAAAAACCATCTGGTGATGATAGATAC 211

Qy  41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
```

```
Db   212 AATCAGTTTCACATAAGGAGATTGAAGCGATTACTTCAAACTGGAAAAGTCCCACTTCT
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   272 GAATTACTTCAAACTGGAAAAGTCCCACTTCT

Qy    61 GluLeuLeuPhe-PheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValPro  80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   272 GAATTACTTCAAACTGGGGCACCACAAATTGCACAGTTGGTCATCTGGTCGATCTTTT 331

Qy    80 ...IleGlnAsnGluPhe-PheAlaProAlaSerLeuLeuLeuPheProAspAlaValPro  98
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   332 GATCCAAAATGAATTTGTTGGGCTCCTGCGAAGGTCTTTTGCTCCCAGATGCTGTTCCC 387
```

RESULT 5
LOCUS   CA538859
CA538859
DEFINITION C0272B01-5N NIA Mouse 7.5-dpc Whole Embryo cDNA Library (Long) Mus
            musculus cDNA clone NIA:C0272B01 IMAGE:30017484 5', mRNA sequence.
ACCESSION   CA538859
VERSION   CA538859.1 GI:25080828
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 503)
AUTHORS   Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Tanaka,T., Lim,M.K.,
            Luo,A. and Ko,M.S.H.
TITLE   Systematic Analyses of NIA Mouse 7.5-dpc Whole Embryo cDNA Library
            (Long)
JOURNAL   Unpublished
COMMENT   Other ESTs: C0272B01-3
            Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: C0272 row: B column: 01
            Seq primer: M13 Reverse
            High quality sequence stop: 503
FEATURES   Location/Qualifiers
     source   1..503
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:C0272B01-5N"
                     /db_xref="taxon:10090"
                     /clone="NIA:C0272B01 IMAGE:30017484"
                     /tissue_type="whole embryo including extraembryonic
                     tissues at 7.5-days postcoitum"
                     /dev_stage="7.5-days postcoitum"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse 7.5-dpc Whole Embryo cDNA Library
                     (Long)"
                     /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA); Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library [Ref. Genome Res.
                     11: 1553-1558 (2001). (PMID: 11544199)]. Total RNAs were
                     extracted from a pool of four embryos at 7.5-days
                     postcoitum. Double-stranded cDNAs were synthesized with an
                     Oligo(dT) primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
                     7 ug of total RNA, treated with T4 DNA polymerase, and
                     purified by ethanol-precipitation. The cDNAs were ligated
                     to Lone-linker Li-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by long-range high fidelity PCR using
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                     The DH10B E. coli host was transformed with the ligation

BASE COUNT   119 a   146 c   138 g   100 t
ORIGIN
```
mixture by the standard chemical method. The average
Insert size is about 2.2 kb. The library was constructed
by Yulan Piao (NIA)."

```
US-10-001-254-6 (1-98) x CA538859 (1-503)

Qy     1 ThrTyrValArgCysLeuAsnValGlyTyrLeuIleLeuIleArgLysLeuSerAspPheIleAspPro  20
             ||||| :::|||   |||| ||||||||||:|||||::::|||||||| ||||||   |||||   |||
Db   192 ACAATACATACGCAACCTTAAATGTGGGGAGGGATCCTTAGGAGCGTCGGAGTTTATTGATCCT 251

Qy    21 GlnGluGlyTrpLysLysLeuAlaValAlaLaIleLysLysProSerGlyAspAspArgTyr  40
             |||||||||||||||   |||| ::::|   |   |  ||||||||||||||||||||
Db   252 CAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAAAAGCCGTCCGGCGACGGACAGATAC 311

Qy    41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   312 AATCAGTTCCATATAAGGAGGATTCGAAGCCTTACTTCAGACCCGGGAAGGAGGCCCCACCTGT 371

Qy    61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   372 GAACTGCTGTTTGACTGGGGCACCACAGAACTGACAGTTGGCGACCTTGTGGATCTACTG 431

Qy    81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuPheProAspAlaValPro  98
             ::::|| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   432 GTCCAGAATGAGCTGTTTGCCCCCGCGCACTCTCTGCTCCCGCGGATGCCGTTCCC 485
```

```
Alignment Scores:
Pred. No.:
Score:                   3.16e-56          Length:          503
                         476.00            Matches:         89
Percent Similarity:      95.92%            Conservative:    5
Best Local Similarity:   90.82%            Mismatches:      4
Query Match:             92.07%            Indels:          0
DB:                      14                Gaps:            0
```

RESULT 6
LOCUS   BQ552228
BQ552228
DEFINITION BQ552228
            H4014C09-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
            H4014C09 5', mRNA sequence.
ACCESSION   BQ552228
VERSION   BQ552228.1 GI:21453114
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 598)
AUTHORS   VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G., Martin
            ,P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T., Kargul,G.J.,
            Luo,A.G., Kelso,J., Hide,W. and Ko,M.S.H.
TITLE   Assembly, verification, and initial annotation of NIA 7.4K mouse
            cDNA clone set
JOURNAL   Genome Res. 12 (12), 1999-2003 (2002)
MEDLINE   22354164
PUBMED   12466305
COMMENT   Contact: Yong Qian
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/NIA_7_4k.html for details.
            Plate: H4014 row: C column: 09
            Seq primer: -21M13 Reverse
            High quality sequence stop: 598
            POLYA=No.
FEATURES   Location/Qualifiers
     source   1..598
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
```

BASE COUNT   150 a   174 c   163 g   111 t
ORIGIN

/db_xref="niaEST:H4014C09-5"
/db_xref="taxon:10090"
/clone="H4014C09"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/clone_lib="NIA Mouse 7.4K cDNA Clone Set"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This
clone is among a rearrayed set of 7,407 clones from more
than 20 cDNA libraries."

US-10-001-254-6 (1-98) x BQ55228 (1-598)

Alignment Scores:
Pred. No.:   4.03e-56
Score: 476.00
Percent Similarity: 95.92%
Percent Identity: 90.82%
Best Local Similarity: 92.07%
Query Match: 13
DB:

Length: 598
Matches: 89
Conservative: 5
Mismatches: 4
Indels: 0
Gaps: 0

Qy   1 ThrThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
         ||||||||||||::|||||||||||||||||||||||||||||||||||||||||||||||
Db 176 ACATACATACCGCAACCTTAAYGTGGGGGATCCTTAGGAAGSCTGTCCGGATTTTATTGATCCT 235

Qy   21 GlnGluGlyTrpLysLeuLysLeuAlaValAlaLalLeLysGlyProSerGlyAspAspArgTyr 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 236 CAAGAGGGGTGGAAGAAATTAGCAGTRAGCTATCAAAAAGCCGTCCGGCGACGACAGATAC 295

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 296 AATCAGTTCCATATAAGGAGATTCGAAGCCTTACTTCAGACCGGGAAGAGCCCCACCTGT 355

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 356 GAACTCGTCGTTTGACTGGGGCACCAACTGCCACAGTTGGCGACCTTGTGGATCTACTG 415

Qy   81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
         ||| ||| |||||||||||:::||||||||||||||||||||||||||||
Db 416 GTCCAGAATTGAGCTGTTTGCCCCGCCACTCTCCTCGTCGCGGATGCCGTTCCC 469

RESULT 7
BB660378
LOCUS   BB660378   610 bp   mRNA   linear   EST 26-OCT-2001
DEFINITION BB660378 RIKEN full-length enriched, 13 days embryo lung Mus
            musculus cDNA clone D430042I21 5', mRNA sequence.
ACCESSION   BB660378
VERSION   BB660378.1 GI:16494199
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 610)
AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL   Unpublished
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp.

URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. . 10 (10), 1617-1630 (2000)
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res..
10 (11), 1757-1771 (2000)
Konno,H. Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara.
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES   Location/Qualifiers
     source   1..610
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="D430042I21"
                     /tissue_type="lung"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     lung"
                     /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCCAGTTTTTTTTTTTTTTTTN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGAGTTCGGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT   149 a   179 c   164 g   118 t
ORIGIN

US-10-001-254-6 (1-98) x BB660378 (1-610)

Alignment Scores:
Pred. No.:   4.15e-56
Score: 476.00
Percent Similarity: 95.92%
Percent Identity: 90.82%
Best Local Similarity: 92.07%
Query Match: 10
DB:

Length: 610
Matches: 89
Conservative: 5
Mismatches: 4
Indels: 0
Gaps: 0

Qy   1 ThrThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
         ||||||||||||::|||||||||||||||||||||||||||||||||||||||||||||||
Db 212 ACATACATACCGCAACCTTAAYGTGGGGGATCCTTAGGAAGSCTGTCCGGATTTTATTGATCCT 271

Qy   21 GlnGluGlyTrpLysLeuLysLeuAlaValAlaLalLeLysGlyProSerGlyAspAspArgTyr 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 272 CAAGAGGGGTGGAAGAAATTAGCAGTRAGCTATCAAAAAGCCGTCCGGCGACGACAGATAC 331

```
RESULT 8

Qy    41  AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer  60
Qy   332  AATCAGTTCCATATAAGGAGATTCGAAGCCTTACTTCAGACCGGGAAGAGCCCCACCTGT 391
Db

Qy    61  GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
Db   392  GAACTGCTGTTTGACTGGGGCACCACGAACTGCACAGTTGGGCACCTTGTGGATCTACTG 451
Db

Qy    81  IleGlnAsnGluPheProAlaProAlaSerLeuLeuLeuProAspAlaValPro  98
Db        :::||  |||||||||||||||||||||||||||||||||||||||||||||
Db   452  GTCCAGATTGAGCTGTTTGCCCCCGCCACTCTCCTGCTGCCGGATGCCGTTCCC 505
```

LOCUS   BB613447   637 bp   mRNA   linear   EST 26-OCT-2001
DEFINITION BB613447 RIKEN full-length enriched, 10 day neonate skin Mus
            musculus cDNA clone 4732482P03 5', mRNA sequence.
ACCESSION   BB613447
VERSION   BB613447.1 GI:16454055
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 637)
AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H. Sasaki
            ,D., Shibata,K., Shinagawa,A., Takahashi,F., Takeda,Y., Tanaka,T.,
            Tagami,M., Tagawa,A., Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL   Unpublished
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T. Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,
            K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) for
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.

FEATURES   Location/Qualifiers
     source   1..637
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"

BASE COUNT   162 a   185 c   173 g   117 t
ORIGIN

RESULT 9

```
Qy     1  ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro  20
Db   193  ACATACATACGGCAACCTTAATGTGGGGGATCCTTAGGAAGCTGTTGGGATTTATTGATCCT 252

Qy    21  GlnGluGlyTrpLysLysLeuAlaValAlaIleLeuIleLeuLysLysProSerGlyTyr  40
Db   253  CAAGAGGGGTGGAAGAAAATTAGCAGTAGCTTATCAAAAAAGCCGTCCGGCGACGACAGATAC 312

Qy    41  AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer  60
Db   313  AATCAGTTCCATATAAGGAGATTCGAAGCCTTACTTCAGACCGGGAAGAGCCCCACCTGT 372

Qy    61  GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu  80
Db   373  GAACTGCTGTTTGACTGGGGCACCACGAACTGCACAGTTGGGCACCTTGTGGATCTACTG 432

Qy    81  IleGlnAsnGluPheProAlaProAlaSerLeuLeuLeuProAspAlaValPro  98
Db        :::||  ||||||||||||||||||||||||||||||||||||||||||||||
Db   433  GTCCAGATTGAGCTGTTTGCCCCCGCCACTCTCCTGCTGCCGGATGCCGTTCCC 486
```

US-10-001-254-6 (1-98) x BB613447 (1-637)

Alignment Scores:
Pred. No.:
Score: 4.41e-56   Length: 637
Percent Similarity: 476.00   Matches: 89
Best Local Similarity: 95.92%   Conservative: 5
Query Match: 90.82%   Mismatches: 4
DB: 92.07%   Indels: 0
                      10   Gaps: 0

LOCUS   BY721552   638 bp   mRNA   linear   EST 17-DEC-2002
DEFINITION BY721552 RIKEN full-length enriched, adult male diencephalon Mus
            musculus cDNA clone 9330209D03 5', mRNA sequence.
ACCESSION   BY721552
VERSION   BY721552.1 GI:27134669
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 638)
AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.

COMMENT
            /db_xref="taxon:10090"
            /clone="4732482P03"
            /sex="mixed"
            /tissue_type="skin"
            /dev_stage="10 days neonate"
            /lab_host="DH10B"
            /clone_lib="RIKEN full-length enriched, 10 day neonate
            skin"
            /note="Site_1: SalI; Site_2: BamHI; cDNA library was
            prepared and sequenced in Mouse Genome Encyclopedia
            Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in
            RIKEN. Division of Experimental Animal Research in Riken
            contributed to prepare mouse tissues. 1st strand cDNA was
            primed with a primer [5'
            GAGAGAGAGAGGATCCAAGAGCTCTGTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse
            transcriptase and subsequently enriched for full-length by
            cap-trapper. cDNA went through one round of normalization
            to Rot = 10.0 and subtraction to Rot = 100.0. Second
            strand cDNA was prepared with the primer adapter of
            sequence [5' GAGAGAGAGATTCGGAGTTAATTAAAATTAAATCCCCCCCCC
            3']. cDNA was cloned into the XhoI and BamHI sites.
            Vector: a modified pBluescript KS(+) after bulk excision
            from Lambda FLC I"
```

```
                            Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
                            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
                            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
                            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
                            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
                            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
                            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
                            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
                            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
                            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
                            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
                            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
                            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
                            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
                            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
                            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
                            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
                            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
                            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
                            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
                            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
                            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
                            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
                            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  PUBMED    12466851
  MEDLINE   22354683
  COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp.
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
            Computational Analysis of Full-length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in Riken.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
  FEATURES             Location/Qualifiers
       source          1..638
                       /organism="Mus musculus"
                       /mol_type="mRNA"
                       /strain="C57BL/6J"
                       /db_xref="taxon:10090"
                       /clone="9330209D03"
                       /sex="male"
                       /tissue_type="diencephalon"
```

```
                            /dev_stage="adult"
                            /lab_host="DH10B"
                            /clone_lib="RIKEN full-length enriched, adult male
                            diencephalon"
                            /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                            prepared and sequenced in Mouse Genome Encyclopedia
                            Project of Genome Exploration Research Group in Riken
                            Genomic Sciences Center and Genome Science Laboratory in
                            RIKEN. Division of Experimental Animal Research in Riken
                            contributed to prepare mouse tissues. 1st strand cDNA was
                            primed with a primer [5'
                            GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTV 3'], cDNA was
                            prepared by using trehalose thermo-activated reverse
                            transcriptase and subsequently enriched for full-length by
                            cap-trapper. cDNA went through one round of normalization
                            to Rot = 10.0 and subtraction to Rot = 185.0. Second
                            strand cDNA was prepared with the primer adapter of
                            sequence [5' GAGAGAGAGATTCGACTTAATTAAATTAATCCCCCCCCCCCC
                            3']. cDNA was cloned into the XhoI and BamHI sites.
                            Vector: a modified pBluescript KS(+) after bulk excision
                            from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                            BamHI"
  BASE COUNT       160 a       189 c       172 g       117 t
  ORIGIN
```

```
  US-10-001-254-6 (1-98) x BY721552 (1-638)

  Alignment Scores:
  Pred. No.:         4.42e-56        Length:          638
  Score:             476.00          Matches:         89
  Percent Similarity: 95.92%         Conservative:    5
  Best Local Similarity: 90.82%      Mismatches:      4
  Query Match:       92.07%          Indels:          0
  DB:                14              Gaps:            0
```

```
  Qy     1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgGlysLeuSerAspPheIleAspPro 20
            ||||||::||||||||||||||||||||||||||::::::||||||||||||||||||||||
  Db   198 ACATACATACGGCAAACCTTAAGTGTGGGAGATCCTTAGGAAGCTGTCGGAGATTTAATTGATCCT 257

  Qy    21 GlnGluGlyTrpLysGlyLeuAlaValAlaIleLeuGlysProSerGlyAspAspArgTyr 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Db   258 CAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAAAAGCCGTCCGCGACCGACAGATAC 317

  Qy    41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Db   318 AATCAGTTCCATATAAGGAGAATTCGAAGCCTTACTTCAGACACGGGAAGAGCCCCACCTGT 377

  Qy    61 GluLeuLeuPheAspIleProGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Db   378 GAAACTGCTGTTTGACTGGGGCACCACAGAAACTGCACAGTTGGCGACCTTGTGGATCTACTG 437

  Qy    81 IleGlnAsnGluPhePheAlaProLaSerLeuLeuLeuProAspAlaValPro 98
            :::||:||| |||||||||||||||||||:||:|||||||||||||||||||||
  Db   438 GTCCAGATTGAGCTGTTTGCCCCCGCCACTCTCCTGCTGCCGGATGCCGTTCCC 491
```

```
  RESULT 10
  LOCUS       BY726858
  DEFINITION  BY726858 RIKEN full-length enriched, adult male corpora
              quadrigemina Mus musculus cDNA clone B230034I03 5', mRNA sequence.
  ACCESSION   BY726858
  VERSION     BY726858.1  GI:27139985
  KEYWORDS    EST.
  SOURCE      Mus musculus (house mouse)
    ORGANISM  Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE   1  (bases 1 to 663)
  AUTHORS     Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
              Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
              , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
              Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
              Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
```

```
  BY726858                663 bp    mRNA        linear    EST 17-DEC-2002
  BY726858.1  GI:27139985
  BY726858
```

Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,
L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,
A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
, Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
,R.D., Tonita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
,Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
,K., Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

JOURNAL   Nature 420, 563-573 (2002)
PUBMED   12466851
MEDLINE   22354683

COMMENT   Analysis of the mouse transcriptome based on functional annotation
          of 60,770 full-length cDNAs

TITLE   The Institute of Physical and Chemical Research (RIKEN)
JOURNAL   Contact: Yoshihide Hayashizaki
          Laboratory for Genome Exploration Research Group, RIKEN Genomic
          Sciences Center(GSC) Yokohama Institute
          1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          Tel: 81-45-503-9222
          Fax: 81-45-503-9216
          Email: genome-res@gsc.riken.go.jp,
          URL: http://genome.gsc.riken.go.jp/
          Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
          ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
          Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
          ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
          Numazaki,R., Ohno,M., Ohsato,N., Saito,N., Sakazume,N., Sano,H.,
          Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
          Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
          Submission
          Computational Analysis of Full-Length Mouse cDNAs Compared with
          Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
          Computer-based methods for the mouse full-length cDNA
          encyclopedia: real-time sequence clustering for construction of a
          nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
          Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new
          genes. Genome Res. 10 (10), 1617-1630 (2000)
          RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer. Genome Res.
          10 (11), 1757-1771 (2000)
          Please visit our web site (http://genome.gsc.riken.go.jp) for
          further details.

FEATURES   Location/Qualifiers
     source   1..663
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="B23034I03"
                     /sex="male"
                     /tissue_type="corpora quadrigemina"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male corpora
                     quadrigemina".

/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT   165 a   190 c   181 g   124 t   3 others
ORIGIN

US-10-001-254-6 (1-98) x BY726858 (1-663)

Alignment Scores:
Pred. No.:
Score: 4.66e-56
Percent Similarity: 476.00
Best Local Similarity: 95.92%
Query Match: 90.82%
                     92.07%
                     14

Length: 663
Matches: 89
Conservative: 5
Mismatches: 4
Indels: 0
Gaps: 0

Qy   1 ThrThrTyrValArgCysLeuAsnAlaValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
          |||||:||| ||||||| ||||| |:||||
Db 185 ACATACATACGCAACCTTAATGTGGGGATCCTTAGGAGAAGCTTGTCGGATTTTATTGATCCT 244

Qy   21 GlnGluGlyTrpLysGlyValAlaValAlaAlaIleLysLysArgProSerGlyAspAspArgTyr 40
          |||||||| ||||||||||||| |||||||||||||||||||||||||||||||||||||
Db 245 CAAGAAGGGTGGAAGAAAATTAGCACAGTTAGTTATCAAAAAGCCGTCCGGCGACGACGAGATAC 304

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrGlyLysSerProThrSer 60
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db 305 AATCAGTTCCATATAAGGAGATTCGAAGCCTTACTTCAGCACCGGGAAGAGCCCACCTGT 364

Qy   61 GlyLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 365 GAACTGCTGTTTGACTGGGGCACCACGAACTGCACAGTTGGCGACCTTGTGGGATCTACTG 424

Qy   81 IleGlnAsnGluPhePheAlaProLaSerLeuLeuLeuProAspAlaValPro 98
          :||||||||||||||||||||||||||:|::|||||||||||||||||||||
Db 425 GTCCAGATTGAGCGTGTTTGCCCCCGCCCCACTCTCCTTGCTGCCGGATGCCGGTTCCC 478

RESULT 11
BB613167
BB613167
LOCUS   BB613167   676 bp   mRNA   linear   EST 26-OCT-2001
DEFINITION RIKEN full-length enriched, 10 day neonate skin Mus
            musculus cDNA clone 4732460109 5', mRNA sequence.
ACCESSION   BB613167
VERSION   BB613167.1 GI:16453871
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 676)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Sasaki
            ,Okazaki,Y., Okido,T., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
  COMMENT   Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp.
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
Watahiki,M., Yoneda,Y., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
,S., Kawai,J., Okazaki,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
Hayashizaki,Y.
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES   Location/Qualifiers
    source   1..676
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4732460I09"
                     /sex="mixed"
                     /tissue_type="skin"
                     /dev_stage="10 days neonate"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 10 day neonate
                     skin"
                     /note="Site 1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'-
                     GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5'-GAGAGAGAGATTCTCGAGTTAATTAAATTAAATCCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I"

BASE COUNT   196 c   172 g   128 t
ORIGIN   180 a

Alignment Scores:
Pred. No.:   4.79e-56
Score: 476.00   Length: 676
Percent Similarity: 95.92%   Matches: 89
Best Local Similarity: 90.82%   Conservative: 5
Query Match: 92.07%   Mismatches: 4
DB: 10   Indels: 0
                                    Gaps: 0

US-10-001-254-6 (1-98) x BB613167 (1-676)

  1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 20
    |||||||||||:::||||||||||||||||:::|||||||||||||||||||||||||||||
  1 ACAATACAATACGCAACCTTAATGTGGGGAGAAGGCTGTCGGGATTTTATTGATCCT 199

 21 GlnGluGlyTrpLysLysLeuAlaValAlaIleAlaIleLysLysProSerGlyAspAspTyr 40
    |||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
200 CAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAAAAGCCTTCCGGGACGACGATAC 259

 41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyTyrSerProThrSer 60
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
260 AATCAGTTCCATATAAGGAGAATTCGAAGCCTTACTTCAGACCGGGATACAGAGCCCACCTGT 319

 61 GluLeuLeuPheAspTrpGlyTyrThrTyrAsnCysThrValGlyValValAspLeuLeu 80
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
320 GAACTGCTGTTTGACTGGGGCACCACCACGAACTGCACAGTTGGGCCACCTTGTGGAATCTACTG 379

 81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuLeuProAlaAspAlaValPro 98
    |||:::|| |||||||||||||||||:::|||||||||||||||||||||||||||
380 GTCCAGAATTGAGCTGTTTGCCCCGGCCAGTCTTCTGCTGCGCGGATGCCGGTTCC 433

RESULT 12
LOCUS   AK020397   1161 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION Mus musculus adult male diencephalon cDNA, RIKEN full-length
            enriched library, clone:9330209D03 product:interleukin-1
            receptor-associated kinase 4 [Mus musculus], full insert sequence.
ACCESSION   AK020397
VERSION   AK020397.2 GI:15919911
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,

```
TITLE       Functional annotation of a full-length mouse cDNA collection
JOURNAL     Nature 409 (6821), 685-690 (2001)
MEDLINE     21085660
PUBMED      11217851
REFERENCE   5
AUTHORS     Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
JOURNAL     Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1161)
AUTHORS     Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,R., Sakai,C., Sakai,Y., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
JOURNAL     Direct Submission
            Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     On Oct 4, 2001 this sequence version replaced gi:12860970.
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAGGATCCGAGGCCATTATGGCCGGG 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through two rounds of normalization to Rot = 20.0 and subtraction
            to Rot = 370.4. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:9330209D03"
                     /db_xref="MGI:1911393"
                     /db_xref="taxon:10090"
                     /clone="9330209D03"
                     /sex="male"
                     /tissue_type="diencephalon"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             183..776
                     /note="unnamed protein product; interleukin-1
                     receptor-associated kinase 4 [Mus musculus]
                     putative"
                     /codon_start=1
                     /protein_id="BAB32090.2"
                     /db_xref="GI:15919912"
                     /db_xref="MGI:1924812"
                     /translation="MNKPLTPSTYIRNIANVGILRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLGTGKSPTCELLFDWGTTNCTVGDLVDLLVQIELFAPATLLLP
                     DAVPQTVKSLPPKEAGNSGTNTRALSGKDQDIRNAYAKTRTQLRATGLISPDNRSVES
                     SDTRFHSFSFHHLKSITNNFDEQPASAGGNRMGEGGF"
```

```
BASE COUNT      317 a    296 c    308 g    240 t
ORIGIN
```

```
US-10-001-254-6 (1-98) x AK020397 (1-1161)

Alignment Scores:
  Pred. No.:          1,02e-55            Length:        1161
  Score:              476.00              Matches:        89
  Percent Similarity: 95.92%              Conservative:    5
  Best Local Similarity: 90.82%           Mismatches:      4
  Query Match:        92.07%              Indels:          0
  DB:                 11                  Gaps:            0

Qy      1 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAspPheIleAspPro   20
          ||||||| ||||:::||||||||||||||||||::||:||||||||||||||||||||||||||
Db    207 ACATACATACGGCAACCTTAATGTGGGAGATCCTTAGGAAGCTAGTCCGGGATTTTATTGATCCT  266

Qy     21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr     40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    267 CAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAAAAACCGTCCGGCGACGACAGATAC  326

Qy     41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer     60
          ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db    327 AATCAGTTCCATATAAGGAGATTCGAAGCCTTACTTCAGACCGGGAAGAGCCCCACCTGT  386

Qy     61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu     80
          ||||||||||||||||||||||||||||||:||||:||||||||||||||||||||||||
Db    387 GAACTGCTGTTTGACTGGGGCACCACGAACTGCACAGTTGGCGACCTTGTGGATCTACTG  446

Qy     81 IleGlnIleGluPhePheAlaProAlaSerLeuLeuLeuProAspAlaValPro     98
          |||:::|||||||||||||||||||||:::||||||||||||||||||||||||
Db    447 GTCCAGATTGAGCTGTTTCCCCCGGCCACTCTCCTGCTCCCGGATGCCGTTCCC  500
```

```
RESULT 13
ACCESSION   AK028837
LOCUS       AK028837    2481 bp    mRNA    linear    HTC 05-DEC-2002
DEFINITION  Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
            library, clone:4732460109 product:interleukin-1 receptor-associated
            kinase 4 [Mus musculus], full insert sequence.
ACCESSION   AK028837
VERSION     AK028837.1  GI:26324783
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Carninci,P. and Hayashizaki,Y.
TITLE       High-efficiency full-length cDNA cloning
JOURNAL     Meth. Enzymol. 303, 19-44 (1999)
MEDLINE     99279253
PUBMED      10349636
REFERENCE   2
AUTHORS     Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE       Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL     Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE     20499374
PUBMED      11042159
REFERENCE   3
AUTHORS     Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
```

REFERENCE   4
  AUTHORS   Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
    TITLE   RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C.A., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
    TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
   MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagami,Y., Takahashi,T., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
    TITLE   Direct Submission
  JOURNAL   Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
  COMMENT   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
            Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
            Nature 420, 563-573 (2002)
            (bases 1 to 2481)
            This cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
  FEATURES   Location/Qualifiers
     source   1..2481
                       /organism="Mus musculus"

                       /mol_type="mRNA"
                       /strain="C57BL/6J"
                       /db_xref="FANTOM_DB:473246010g"
                       /db_xref="taxon:10090"
                       /clone="4732460109"
                       /tissue_type="skin"
                       /clone_lib="RIKEN full-length enriched mouse cDNA library"
                       /dev_stage="10 days neonate"
                       /note="unnamed protein product; interleukin-1
                       receptor-associated kinase 4 [Mus musculus]
     CDS   116..1495
                       /codon_start=1
                       /protein_id="BAC26146.1"
                       /db_xref="GI:26324784"
                       /translation="MNKPLTPSTYIRNLNVGILRKLSDFIDPQEGWKKLAVAIKPSG
                       DDRYNQFHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLIVQIELFAPATLLLP
                       DAVPQTVKSLPPREAATVAQTHGPCQEKDRTSVMPKLEHSCEPPDSSSPDNRSVES
                       SDTRFHSFSPHELKSITNNFDEQPASAGNRMGEGGGFGVVYKGVVNTIVAKVKLGAM
                       VEISTEELKQQFDQEIKWMATCQHENLVELLGFSSDSDNLCLVYAYMPNGSLLDRLSC
                       LDGTPPLSWHTRCKVAQGTANGIRFLHENHIHRDIKSANILLDKDFTAKISDFGLAR
                       ASARLAQTWMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLELITGLAAVDENRE
                       POLIDIKEEIEDEEKTIEDYTDEKMSDADPASVEAMYSAASQCLHEKKQRRPDIAKV
                       QQLLQEMSA"
ORIGIN
BASE COUNT   650 a   592 c   634 g   605 t

US-10-001-254-6 (1-98) x AK028837 (1-2481)

Alignment Scores:
  Pred. No.:
  Score: 2.98e-55   Length: 2481
  Percent Similarity: 476.00   Matches: 89
  Best Local Similarity: 95.92%   Conservative: 5
  Query Match: 90.82%   Mismatches: 4
  DB: 92.07%   Indels: 0
                              11   Gaps: 0

Qy   1 ThrThrTyrValArgCysLeuAsnValGlyLeuLeuArgLysLeuSerAspPheIleAspPro 20
            ||||||||:::|| ||||||||:::||::|||||||:||:||||||||||||||||||||
Db   140 ACATACATACCGCAACCTTAAYGTGGGGATCCTTAGGAAAGCTGTCGGATTTTATTGATCCT 199

Qy   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysProSerGlyAspAspArgTyr 40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   200 CAAGAGGGGTGGAAGAAATTAGCAGTAGCTATCAAAAAAGCCGGTCCGGACGACAGATAC 259

Qy   41 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSerProThrSer 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   260 AATCAGTTCCATATAAGGAGAATTCGAAGCCTTACTTCAGACCGGGAAGAGCCCACTGT 319

Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuIleVal 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   320 GAACTGCTGTTTGACTGGGGCACCACGAACTGCACAGTTGGCGACCTTGTGGATCTACTG 379

Qy   81 IleGluGlnAsnGluPheHisAlaProAlaSerLeuLeuLeuProAspAlaValPro 98
            ::|||| || ||||||||||||:::||||||||||||||||||||||||||||||
Db   380 GTCCAGATTGAGCGTGTTTGCCCCCGGCCACTCTCCTCCTCGCTGCCGGATCCGGTTCCC 433

REFERENCE
  AUTHORS   Carninci,P. and Hayashizaki,Y.
    TITLE   High-efficiency full-length cDNA cloning

RESULT 14
AK029028
LOCUS   AK029028   2810 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
            library, clone:4732482P03 product:interleukin-1 receptor-associated
            kinase 4 [Mus musculus], full insert sequence.
ACCESSION   AK029028
VERSION   AK029028.1 GI:26081132
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
    TITLE   High-efficiency full-length cDNA cloning

JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED   10349636
REFERENCE 2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
           Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
           prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE 3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
           Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
           Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
           Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
           Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
           Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
           Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
           sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE 4
AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
           Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
           Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
           Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
           Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
           Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
           Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
           Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
           Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
           Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
           Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
           Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
           Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
           Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
           Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
           Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
           Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
           Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
           and Hayashizaki,Y.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
MEDLINE   21085660
PUBMED   11217851
REFERENCE 5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
           Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
           of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2810)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
           Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
           Hayashida,K., Hayatsu,N., Hiramoto,K., Hiramoto,T., Hiraoka,T.,
           Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
           Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
           Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
           Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
           Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
           Sano,H., Sasaki,D., Shibata,K., Shiraki,T.,
           Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
           Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
           Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,

COMMENT   URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
           CDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in Riken
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues.
           Please visit our web site for further details.
           URL:http://genome.gsc.riken.go.jp/
           URL:http://fantom.gsc.riken.go.jp/.

FEATURES   Location/Qualifiers
     source   1..2810
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:4732482P03"
                     /db_xref="taxon:10090"
                     /clone="4732482P03"
                     /tissue_type="skin"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     misc_feature   176..1556
                     /note="interleukin-1 receptor-associated kinase 4 [Mus
                     musculus]
                     putative"

BASE COUNT   738 a   677 c   715 g   680 t
ORIGIN

US-10-001-254-6 (1-98) x AK029028 (1-2810)

Alignment Scores:
Pred. No.:
Score: 3.54e-55
Percent Similarity: 476.00   Length: 2810
Best Local Similarity: 95.92%   Matches: 89
Query Match: 90.82%   Conservative: 5
DB: 92.07%   Mismatches: 4
                        11   Indels: 0
                                      Gaps: 0

    1 ThrTyrValArgCysLeuLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro   20
      |||||::||| |||||||||::|||||||||||||||||||||||| |||||||||||||||||
    1 ACATACTACGCAACCTTAATGTGGGGGATCCTTAGGAAGCTGTCGGATTTTATTGATCCT   259

   21 GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysArgProSerGlyAspAspArgTyr   40
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  260 CAAGAAGAGGGTGGAAGAAAATTAGCAGTTAGCTATCAAAAAGCCGTCCGGCGACGACAGATAC   319

   41 AsnGlnPheHisIleArgArgPheGluAlaIleLeuLeuGlnThrGlyLysSerProThrSer   60
      ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
  320 AATCAGTTCCATATAAGGAGATTCGAAGCCTTACTTACTTCAGACCGGGAAGAGCCCACTGT   379

   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeuLeu   80
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  380 GAACTGCTGTTTGACTGGGGCACCACGAACTGCACAGTTGGCGACCTTGTGGATCTACTG   439

   81 IleGlnAsnGluPheAlaProAlaSerLeuLeuLeuProAspAlaValPro   98
      ::|||||||| |||:::||||||||||||||||||||||
  440 GTCCAGATTGAGCTGTTGCCCCCCCCACTCTCCTGCTGCGCGGATGCGTTCCC   493

RESULT 15
BX522921
ID   BX522921   standard; RNA; EST; 575 BP.
XX
AC   BX522921;
XX
SV   BX522921.1
XX
DT   27-MAY-2003 (Rel. 75, Created)
DT   27-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   RZPD Mus musculus cDNA clone IMAGp998I1225504 = IMAGE:2225445 5' EST.
XX
KW   EST; expressed sequence tag.
XX

```
OS   Mus musculus (house mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-575
RA   Heil O., Ebert L., Neubert P., Peters M., Radelof U., Schneider D.,
RA   Korn B.;
RT   ;
RL   Submitted (28-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Neuenheimer
RL   Feld 580, D-69120 Heidelberg, Germany
XX
RL   RZPD; IMAGp998L225504.
RL   RZPDlib; I.M.A.G.E. cDNA Clone Collection;
CC   Mouse UnigeneSet - RZPD2 (RZPDLIB No.981)
CC   http://www.rzpd.de/cgi-bin/showLib.pl.cgi/response?libNo=981
CC   Contact: Ina Rolfs
CC   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
CC   Heubnerweg 6, D-14059 Berlin, Germany
CC   Tel: +49 30 32639 101
CC   Fax: +49 30 32639 111
CC   www.rzpd.de
CC   This clone is available royalty-free from RZPD;
CC   contact RZPD (clone@rzpd.de) for further information.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..575
FT                   /db_xref="taxon:10090"
FT                   /note="1st strand cDNA was primed with an oligo(dT) primer
FT                   [ATGTGGCCGCTTTTTTTTTTTTTTTTTT]; double-stranded cDNA was
FT                   ligated to a DraIII adaptor [TGTTGGCCCTACTGG], digested and
FT                   cloned into distinct DraIII sites of the pME18S-FL3 vector
FT                   (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used
FT                   to isolate the cDNA insert. Size selection was performed to
FT                   exclude fragments <1.5kb. Library constructed by Dr. Sumio
FT                   Sugano (University of Tokyo Institute of Medical Science).
FT                   Custom primers for sequencing: 5' end primer
FT                   CGACCTGCAGCTCGACGACA. REFERENCES: Suzuki, Y., Yoshitomo,
FT                   K., Maruyama, K., Suyama, A., and Sugano, S. Construction
FT                   and characterization of a full length-enriched and a 5' end
FT                   enriched cDNA library. Gene 200, 149-156, 1997. Sasaki, Z.,
FT                   Suzuki, Y., Watanabe, M., Imai, J., Shibui, A., Yoshida, Z.,
FT                   K., Hata. H., Yamaguchi, R., Tateyama, S., and Sugano, S.
FT                   Construction of mouse full length-enriched cDNA libraries
FT                   by oligo-capping. DNA Research, submitted."
FT                   /organism="Mus musculus"
FT                   /clone="IMAGp998L225504"
FT                   /clone_lib="Sugano mouse embryo mewa"
FT                   /dev_stage="embryo, 14 dpc"
FT                   /lab_host="DH10B"
XX
SQ   Sequence 575 BP; 132 A; 172 C; 155 G; 114 T; 2 other;
     Seq primer: sugF, Primer sequence: CTTCTGCTCTAAAAGCTGCG
```

Db   300 CAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAAAAGCCGTCCCGGGCGACAGATAC 359
Qy   41 AsnGlnPheHisIleArgPheGluAlaLeuLeuLeuGlnThrGlyLysSerProThrSer 60
Db   360 AATCAGTTCCATATAAGGAGATTCGAAGCCTTACTTCAGACCCGGAAGAGCCCCAACCTGT 419
Qy   61 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuValAspLeuLeu 80
Db   420 GAACTGCTTGTTTGACTGGGGCACCACCAGAACTGCACAGTTGGCGGATCTGGGATCTACTG 479
Qy   81 IleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuLeuProAspAlaValPro 98
Db   480 GTCCAGATTGAGCTGTTTGGCCCCGGCCCACTCTCCTGCTGGCCGGAAGCCGTTCCC 533

```
US-10-001-254-6 (1-98) x BX522921 (1-575)

Alignment Scores:
Pred. No.:                     1.39e-55
Score:                         472.00
Percent Similarity:            95.92%
Best Local Similarity:         89.80%
Query Match:                   91.30%
DB:                            4

Qy    1 ThrTyrValArgCysLeuAsnValGlyLeuIleArgGlyLeuSerAspPheIleAspPro 20
        |||||::||::||   ||||::::|||||   |||||||||||||||||||||||
Db  240 ACATACATAGGCAACCTTAATGTGGGGGGATCCTTAGGAAAGCTGTCGGATTTTATTGATCCT 299

Qy   21 GlnGluGlyTrpGlyLysProSerGlyAspAspArgTyr 40
        |||||||||||||||||||||||||||||||||||
```

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 00:40:22 ; Search time 551.702 Seconds
                 (without alignments)
                 2250.748 Million cell updates/sec

Title: US-10-001-254-16
Perfect score: 2386
Sequence: 1 MNKPITPSTYVRCLNVGLIR............KNKRPDIKKVQQLLQEMTAS 460

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext   0.5
                 Ygapop 10.0 , Ygapext   0.5
                 Fgapop 6.0 , Fgapext   7.0
                 Delop   6.0 , Delext   7.0

Searched: 252756 seqs, 134719017 residues

Total number of hits satisfying chosen parameters: 5105512

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn21/USPTO spool/US10001254/runat 16012004 152423 19704/app query.fasta_1.1109
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US10001254 @CGN 1 1 0 @runat 16012004 152423 19704 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   N_Geneseq_19Jun03:*
        1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
        2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
        3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
        4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
        5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
        6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
        7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
        8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
        9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
       10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
       11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
       12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
       13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
       14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
       15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
       16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
       17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
       18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
       19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
       20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
       21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
       22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
       23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
       24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
       25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

---

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID		Description
1	2386	100.0	1383	24	AAD40079	Human IRAK4 gene #
2	2382	99.8	1383	22	AAD10197	Human interleukin-
3	2362	99.0	2817	23	AAD40085	Human IRAK4 gene #
4	2017	84.5	1542	22	AAD10198	Mouse interleukin-
5	1787.5	74.9	2213	22	AAH13798	Human cDNA sequenc
6	1396	58.5	1668	23	AAY76805	DNA encoding novel
7	1169	49.0	833	22	AAA09319	Human cancer assoc
8	972	40.7	1493	22	AAS27214	cDNA encoding nove
9	972	40.7	1493	23	ABK44003	DNA encoding novel
10	972	40.7	1618	23	ABK43727	DNA encoding novel
11	959	40.2	1181	24	ABL90822	Human polynucleoti
12	883	37.0	539	22	AAR22618	Human breast cance
13	883	37.0	586	22	AAP22511	Human breast cance
14	854	35.8	577	22	AAH09161	Human cDNA clone (
15	779	32.6	483	19	AAV63193	cDNA from clone eh
16	779	32.6	483	23	ABQ52043	Human polynucleoti
17	691	29.0	501	24	AAS61608	Lung small cell ca
18	613.5	25.7	415	23	AAS76802	DNA encoding novel
19	517	21.7	294	18	AAD40074	Human IRAK4 DD (de
20	515.5	21.6	3590	24	AATE62551	Interleukin-1 rece
21	515.5	21.6	3590	24	ABJ34976	Interleukin-1 rece
22	515.5	21.6	3590	24	ABK83909	Human gene express
23	515.5	21.6	3590	24	ABN95724	Human cDNA differe
24	515.5	21.6	3590	24	ABL62666	Gene #2222 used to
25	515.5	21.6	3590	25	ACA56495	Colon adenocarcino
26	515	21.6	2622	21	AAG49420	Human signalling p
27	508.5	21.3	2361	24	AB213868	Arabidopsis thalia
28	506	21.2	1899	24	AB212994	Arabidopsis thalia
29	505.5	21.2	1385	24	AAC42456	Arabidopsis thalia
30	504	21.1	1810	24	AAA11352	Arabidopsis thalia
31	501.5	21.0	1794	24	AAL45898	A thaliana recepto
32	501	21.0	1555	24	AAC37541	Arabidopsis thalia
33	499	20.9	1281	24	AB212819	Arabidopsis thalia
34	498.5	20.9	2607	24	AB213464	Arabidopsis thalia
35	496	20.8	1410	21	AAG42305	Arabidopsis thalia
36	496	20.8	1633	21	AA353365	Arabidopsis thalia
37	495.5	20.8	1628	21	AAC47551	Arabidopsis thalia
38	494	20.7	1389	24	AB212535	Arabidopsis thalia
39	493	20.7	1684	21	AAC41148	Arabidopsis thalia
40	489.5	20.5	1674	24	AAC39626	Arabidopsis thalia
41	485	20.3	1239	24	AB213385	Arabidopsis thalia
42	485	20.3	1703	21	AAC46902	Zea mays DNA fragm
43	484	20.3	1629	24	AAL45902	A thaliana recepto
44	480.5	20.1	2640	24	AB213465	Arabidopsis thalia
45	479	20.1	1275	24	AB214172	Arabidopsis thalia

RESULT 1
AAD40079

ID   AAD40079 standard; DNA; 1383 BP.
XX
AC   AAD40079;
XX
DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 gene #1.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
OS   Homo sapiens.

ALIGNMENTS

```
XX
FH   Key              Location/Qualifiers
FT   CDS              1..1383
FT                    /tag= a
FT                    /product= "Human IRAK4"
XX
PN   WO20024068O-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US44844.
XX
PR   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24859.
XX
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection and other diseases
PT   -
XX
PS   Claim 19; Page 180-182; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis, C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 gene.
XX
SQ   Sequence 1383 BP; 463 A; 243 C; 283 G; 394 T; 0 other;
```

```
Alignment Scores:

Pred. No.:
Score:                    6.59e-234
Percent Similarity:       2386.00
Best Local Similarity:    100.00%
Query Match:              100.00%
DB:                       24

US-10-001-254-16 (1-460) x AAD40079 (1-1383)
```

```
Qy          1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg    20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAACATGTTGGACTAATTAGG   60
```

```
Qy         21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLeuLeuAlaLeuAlaIleLeuLys    40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA   120
```

```
Qy         41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleAspArgPheGluAlaLeuLeu    60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAGGACGATTTGAAGCATTACTT   180
```

```
Qy         61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr    80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAAACTGGAAAAAGTCCACTTCTGAATTACTGTTTGACTGGGGACACCACAAATTGCACA   240
```

```
Qy         81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeu   100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGGGATCTTGTGGATCTTTTGATTCAAAATGAATTTTTTGCTCCTGCCAGTCTTTTG   300
```

```
Qy        101 LeuProAspAlaValAlaProLysThrLeuProSerLysGluAlaIleThr             120
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTCCCAGATGCTGTTGCCCCAAAAACTCTAAATCACTACCTTCTAAAGGAAGCTATAACAGTT   360
```

```
Qy        121 GlnGlnIleLysMetProMetCysGluAspLysAspLysArgThrLeuMetThrProValGlnAsn   140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGCAAAAAACAGATGCCTTTGTGTGTGAAGACAAAGATAAAAGGACATTGATGACACCTGTGCAGAAT   420
```

```
Qy        141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal   160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTGAACAAAGCTATATGCCACCTGACTCCTCAAGTCCAGAAAATAAAAGTTTAGAAGTT   480
```

```
Qy        161 SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe   180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGTGATACACGTTTTCACAGTTTTTCATTTTATGAAYTGAAGAATGTCACAAATAACTTT   540
```

```
Qy        181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal   200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATGAACGACCCATTTCTGTTGGTGGTAATAAAATGGGAGAGGGAGGATTTGGAGTTGTA   600
```

```
Qy        201 TyrLysGlyTyrValAsnAsnThrThrValAlaValLysLysLeuAlaAlaMetValAsp   220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TATAAAGGCTACGTAAATAACACAACTGTGGCAGTGAAGAAGCTTGCAGCAATGGTTGAC   660
```

```
Qy        221 IleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysValMetAlaLysCys   240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ATTACTACTGAAGAACTGAAACAGCAGTTTGATCAAGAAATAAAAGTAATGGCAAAGTGT   720
```

```
Qy        241 GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu   260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CAACACGAAAACTTAGTAGAACTACTAGGATTTTCTAGTGATGGAGATGACCTCTGCTTA   780
```

```
Qy        261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr   280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTATATGTTTACATGCCTAATGGTTCAATTGCTTGATAGAGACTCTCGTGCTTGGGATGGTACT   840
```

```
Qy        281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn   300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCACCACTTTCTTGGCACATGAGATGCAAGATTGCTCAGGGTGCAGCTAATGGCATCAAT   900
```

```
Qy        301 PheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp   320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTTCTACATGAAAATCATCATATTCATAGAGATATAAAAGTCAAATATCTTACTCGGAT   960
```

```
Qy        321 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla   340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GAAGCTTTACTGCTAAAAATATCTGACTTTGGCCTTGCACGGGCTTCAGAAAAGTTTGCC   1020
```

```
Qy        341 GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu   360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CAGACAGTTATGACTAGCAGAATTGTGGGAACAACAGCTTATATGGCACCAGAAGCTTTG   1080
```

```
Qy        361 ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle   380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGTGGAGAAATAACACCCAAATCTGATATTTACTCTTTTGGTGTTGTGCTTTTGGAAATA   1140
```

```
Qy        381 IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAspIleLys   400
              ||||||||||||||||||||||||||||
Db       1141 ...
```

```
Db  1141 ATAACTGGACTTCCAGCTGTGGATGACAGACCGTGAACCTGAACCTCAGTTATATTGCTAGATATTAA 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   401 GluGluIleGluAspGluGlyLysThrIleGluAspTyrIleAspProLysLysMetAsnAsp 420
Db  1201 GAAGAAATTGAAGATGAAGAAAAGACAATTGAAGATTATATTGATAAAAAGATGAATGAT 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   421 AlaAspSerThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys 440
Db  1261 GCTGATTCCACTTCAGTTGAAGCTATGTACTCTGTTGCTAGTCAATGTCTTGCATGAAAAG 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   441 LysAsnLysArgProAspIleLysLysValGlnGlnLeuLeuGlnGluMetThrAlaSer 460
Db  1321 AAAAATAAGAGACCAGACATTAAGAAGGTTCAACAGCTGCTGCAAGAGATGACAGCTTCT 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       AAAAATAAGAGACCAGACATTAAGAAGGTTCAACAGCTGCTGCAAGAGATGACAGCTTCT 1380
```

RESULT 2

```
ID   AAD10197 standard; cDNA; 1383 BP.
XX
AC   AAD10197;
XX
DT   24-SEP-2001 (first entry)
XX
DE   Human interleukin-1 (IL-1) receptor-associated kinase (IRAK)-4 cDNA.
XX
KW   Human; interleukin-1 receptor associated kinase-4; IRAK-4; cytostatic;
KW   IL; antibacterial; antiinflammatory; ophthalmological; vasotropic; OPD;
KW   chronic obstructive pulmonary disease; neuroprotective; chronic cough;
KW   adult respiratory disease syndrome; pulmonary fibrosis; asthma; ARDS;
KW   interstitial lung disease; allergic rhinitis; transplant rejection;
KW   autoimmune disease; rheumatoid arthritis; systemic lupus erythematosus;
KW   multiple sclerosis; diabetes; cancer; solid tumour; lymphoma; stroke;
KW   cardiovascular disease; atherosclerosis; neurodegenerative disease;
KW   sepsis; osteoarthritis; rash; osteoporosis; psoriasis; dermatitis;
KW   inflammatory bowel disease; gout; Crohn's disease; ulcerative colitis;
KW   Behcet's syndrome; ankylosing spondylitis; IL-1 receptor/Toll receptor;
KW   sarcoidosis; transgenic animal; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..1383
FT                  /tag= a
FT                  /product= "Human IRAK-4"
XX
SQ   Sequence 1383 BP; 463 A; 244 C; 283 G; 393 T; 0 other;
```

Alignment Scores:

```
Pred. No.:                 1.69e-233
Score:                     2382.00
Percent Similarity:        99.78%
Best Local Similarity:     99.78%
Query Match:               22
```

Length:	1383	
Matches:	459	
Conservative:	0	
Mismatches:	1	
Indels:	0	
Gaps:	0	

DB: US-10-001-254-16 (1-460) x AAD10197 (1-1383)

```
Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 60

Qy    21 LysLeuSerAspAspPheIleAlaAspProGlnGluGlyTrpLysLysLeuAlaValAlaLeuLys 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGCTGTCAGATGATTTTATTGATGCCCAAGAGGAATGGAAGAAGTTAGCTGTAGTATTAAA 120

Qy    41 LysProSerGlyAspAspAspAspAspArgTyrAsnGlnPheHisSleArgArgPheGluAlaLeuLeu 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AAACCATCTGGTGATGATGATAGATACAATCAGTTTCACATAAGGAGAATTGAAGCATTACTT 180

Qy    61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAAACTGGAAAAAGTCCCACTTCTGAATTACGTGTTGACTGGGGCACCACAAATTGCACA 240

Qy    81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPheAsnAlaProAlaSerLeuLeu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCTGGTGATCTTGTGGATCTTTTGATCCAAAATGAATTTAATGCTCCTGCGAGTCTTTTG 300

Qy   101 LeuProAspAlaValProLysThrAlaAlaAsnThrLeuProSerIleLysSerSerLeuThrVal 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTCCCAGATGCTGTTCCAAAACTGCTAATACACTACCTTCTAAAGAAGAGCTATAACAGTT 360

Qy   121 GlnGlyLysGlnMetProPheCysAspLysAspSerSerProGluAsnLysProValGlnAsn 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CAGGCAAAAACAAATGCCTTTTCTGTGACAAAGACAGGAGCAATTGAATGACACTGTGCAGAAT 420

Qy   141 LeuGluGlnSerTyrMetProArgProAspAspSerGluAspSerLysSerLeuGluVal 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CTTGAACAAAGTTATATGCCACCTGACTCCTCAGATTCAGAAGTTTAGAAGGTT 480

Qy   161 SerAspThrArgPheHisSerPhePheTyrGluLeuLysValThrAsnAsnPhe 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AGTGATACAACGTTTTCACAGTTTTTCATTTATGAAGTTAAAGTCACAAATAACTTT 540

Qy   181 AspGluArgProIleSerValGlyGlyAsnIleMetGlyGluGlyGlyPheGlyValVal 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GATGAACGACCACCATTTCTGTTGGTGGGTAATATAATGGGAGAGGGAGGATTTGGTGTA 600
```

The present sequence is a cDNA encoding human interleukin (IL)-1 receptor associated kinase (IRAK)-4. IRAK associate with activated IL-1, IL-18 and other receptors and act to transduce signals originating from the activated receptors, ultimately leading to a variety of downstream effects such as nuclear factor (NF)-kappaB activation. The IRAK-4 inhibitors are useful for treating inflammatory diseases such as pulmonary diseases and diseases of the airway (e.g., adult respiratory

disease syndrome (ARDS), chronic obstructive pulmonary disease (OPD), pulmonary fibrosis, interstitial lung disease, asthma, chronic cough or allergic rhinitis), transplant rejection, autoimmune diseases (e.g., rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis or diabetes), cancer (e.g., solid tumours, skin cancer or lymphoma), cardiovascular diseases (e.g., stroke and atherosclerosis), diseases of the central nervous system (e.g., neurodegenerative disease), CD14 mediated sepsis, non-CD14 mediated sepsis, osteoarthritis, osteoporosis, psoriasis, diseases of the skin (e.g., rash, contact dermatitis, atopic dermatitis), inflammatory bowel diseases, ankylosing spondylitis, gout, sarcoidosis and, ophthalmic diseases and conditions. The inhibitors of IRAK-4 activity or expression are used to inhibit signal transduction resulting from the activation of an interleukin-1 receptor (IL-IR)/Toll receptor in a cell. They also inhibit the activation of a transcription factor that activates NFkappaB in the cell. IRAK-4 is used to create a nonhuman transgenic animal which is useful for testing the function of IRAK-4 in vivo, to generate models for the study of inflammatory disorders and conditions and for the development of potential treatments for IRAK-4-related inflammatory diseases and conditions. IRAK-4 sequences are also used in gene therapy and in antisense therapy.

```
Qy   441 LysAsnLysArgProAspIleLysLysValGlnGlnLeuLeuGlnGluMetThrAlaSer 460
Db  1261 GCTGATTCCACTTCAGTTGAAGCTATGTACTCTGTTGCTAGTCAATGTCTTGCATGAAAAG 1320
```

```
AAD10197
AAD10197
AC   AAD10197;
```

CC   The present sequence is a cDNA encoding human interleukin (IL)-1 receptor
CC   associated kinase (IRAK)-4.

Db   WO20015164l-A1.
PN   WO20015164l-A1.
PD   19-JUL-2001.
XX   12-JAN-2001; 2001WO-US01171.
PF   12-JAN-2001; 2001WO-US01171.
PR   13-JAN-2000; 2000US-0176395.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Wesche H, Li S;
XX
DR   WPI; 2001-451860/48.
DR   P-PSDB; AAE05398.
XX
PS   Claim 7; Fig 2; 89pp; English.

PT   Novel human interleukin-1 receptor associated kinase polypeptide,
PT   useful for identifying modulators of the polypeptide for treating gout,
PT   asthma, allergic rhinitis, multiple sclerosis and skin cancer -

```
Qy  201 TyrLysGlyTyrValAsnAsnThrThrValValAlaValIleLysLeuLeuAlaAlaMetValAsp 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TATAAAGGCTACGTAAATAACAACACTGTGGCAGTGGAAGAGCTTCTGGCAGCAATGGTTGAC 660

Qy  221 IleThrThrGluGlyGluLeuLysGlnPheAspGlnIleLeuLysValMetAlaLysCys 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 ATTACTACTGAAGGAGAACTGAAACAGCAGTTTGATCAAGAAATAAAAGTAATGGCAAAGTGT 720

Qy  241 GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu 260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 CAACATGAAAACTTAGTAGAACTACTTGGTTTCTCAAGTGATGGAGATGACCTGTGCTTA 780

Qy  261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr 280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GTATATGTTTACATGCCTAATGGTTCATTGCTAGACAGACTCTCTTGCTTGGATGGTACT 840

Qy  281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 CCACCACTTTCTTGGCACATGAGATGCAAGATTGCTCAGGGTGCAGCTAATGGCATCAAT 900

Qy  301 PheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp 320
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 TTTCTACATGAAAATCATCATATTCATAGAGATATTAAAAGTGCAAATATCTTACTGGAT 960

Qy  321 GluAlaLeuThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla 340
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 GAAGCTTTACTGCTAAAAATATCTGACTTTGGCCTTGCACGGGCCTCTGAGAAGTTTGCC 1020

Qy  341 GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 CAGACAGTCATGACTAGCAGAATTGTGGGAACAACAGCTTATATGGCACCAGAAGCTTTG 1080

Qy  361 ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle 380
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 CGTGGAGAAATAACACCAAAATCTGATATTTACAGCTTTGGTGTGGTTTTACTAGAAATA 1140

Qy  381 IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAspIleLys 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 ATAACTGGACTTCCAGCTGTGGATGAACACCGTGAACCTCAGTTATTGCTAGATATTAAA 1200

Qy  401 GluGluIleGluGluTrpLysArgAspLysLeuLysIleIleValLysLysAspAsnAsp 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 GAAGAAATTGAAGAATGGAAGAGAGACAAATTGAAGATTATTGTAAAAAAGGATAATGAT 1260

Qy  421 AlaAspProProSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys 440
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 GCTGATCCACCTTCAGTTGAAGCCATGTACTCTGTTGCATCTCAGTGCTTGCATGAAAAG 1320

Qy  441 LysAsnLysArgProAspIleLysLysValGlnGlnLeuLeuGlnGluMetThrAlaSer 460
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 AAAAATAAGAGACCAGACATTAAGAAAGTTCAACAGCTTCTGCAAGAGATGACAGCTTCT 1380
```

RESULT 3

AAD40085

ID   AAD40085   standard; DNA; 2817 BP.
XX
AC   AAD40085;
XX
DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 gene #2.
XX
OS   Homo sapiens.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
XX
FT   Key   Location/Qualifiers
FT   CDS   50..1432
FT   /*tag= a
FT   /product= "Human IRAK4"
XX
PN   WO020240680-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US49844.
XX
PR   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-ljewen F;
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAB24865.
XX
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection and other diseases
PT   -
XX
PS   Claim 19; Page 194-196; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis, C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 gene.
XX
SQ   Sequence 2817 BP; 912 A; 547 C; 586 G; 772 T; 0 other;

US-10-001-254-16 (1-460) x AAD40085 (1-2817)

Alignment Scores:
Pred. No.:
Score: 5.07e-231
                        2362.00
Percent Similarity: 99.13%   Length: 2817
Best Local Similarity: 99.13%   Matches: 456
Query Match: 98.99%   Conservative: 0
                                        Mismatches: 4
                                        Indels: 0
DB: Gaps: 0
                                        24

```
Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLysLeuAsnValGlyLeuIleArg 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCCAAATGTTGGACTAATTAGG 109

Qy   21 LysLeuSerAspPheIleAspProGlnGlyTrpLysLysLeuAlaValAlaAlaIleLys 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  110 AAGCTGTCAGATTTTATTGATCCTCAAGGATGGAAGAAGTTAGCTGTAGCTGCTATTAAA 169
```

```
Qy   41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu   60
         ||||||||||||||||||||||||||||||||||||||||
Db  170  AAACCATCTGGTGATAGATAGACATCAATCAGTTCACATAAGGAGATTTGAAGCATTACTT  229

Qy   61  GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr   80
         |||||||||||||||||||||||||||||||||||||||||||
Db  230  CAAACTGGAAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCACCACAAATTGCACA  289

Qy   81  ValGlyAspLeuValAspAspLeuIleGlnAsnGluPhePheAlaSerLeuLeu  100
         |||||||||||||||||||||||||||||||||||||
Db  290  GTTGGTGAGATCTTTTGATCCAAAATGAATTTTTTGCTTCCTGCGAGTCTTTTG  349

Qy  101  LeuProAspAlaValProLysPheThrAlaAsnThrLeuProSerLysLysGluAlaIleThrVal  120
         ||||||||||||||||||||||||||||||||||||||||
Db  350  CTCCCAGATGCTGTTCCCAAAACTGCTAATACATCACCTTCTAAAGAAGCTATAACAGTT  409

Qy  121  GlnGlyLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn  140
         |||||||||||||||||||||||||||||||||||||||||||
Db  410  CAGCAAAAACAGATGCCTTTCTGTGACAAAGAGAGGACAATTGATGACACCTGTGCAGAAT  469

Qy  141  LeuGluGlnSerGlnTyrMetProProAspSerSerSerProGluAsnLysSerLeuGlyVal  160
         |||||||||||||||||||||||||||||||||||||
Db  470  CTTGAACAGAGTCAATATATGCCACCTGACTCCTCAAGTCCCAGAAAATAAAAGTTTAGGAGTT  529

Qy  161  SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe  180
         ||||||||||||||||||||||||||||||||||||||
Db  530  AGTGATACACGTTTTCACAGTTTTTCATTTTATGAAATTGAAGAATGTCACAAATAACTTT  589

Qy  181  AspGluArgProIleSerValGlyLysValAsnLysSerMetGlyValGlyPheGlyValVal  200
         |||||||||||||||||||||||||||||||||
Db  590  GATGAACGACCCATTTCTGTTGGTGGTAATAAAATGGGAGGAGGATTTGGAGTTGTA  649

Qy  201  TyrLysGlyTyrValAlaAsnAsnThrThrValAlaAlaValLysLysLeuAlaAlaMetValAsp  220
         |||||||||||||||||||||||||||||||||||||
Db  650  TATAAAGGCTACGTAAATAACACAACTGTGGCAGCTGTGAAGAAGCTTGCAGCAATGGTTGAC  709

Qy  221  IleThrThrArgGluGlnLeuLysGlnGlnIleLysGluValMetAlaLysCys  240
         ||||||||||||||||||||||||||||||||||
Db  710  ATTACTACTCGAGAACAGCTTAAACAGCAAATAAAAGAAGTTATGGCAAAGTGT  769

Qy  241  GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu  260
         |||||||||||||||||||||||||||||||||||||||
Db  770  CAACATGAAAACTTAGTAGAACTACTTGGGTTTTCAAGTGATGGAGATGACCTCTGCTTA  829

Qy  261  ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuGlyThr  280
         |||||||||||||||||||||||||||||||||||
Db  830  GTATATGTTTACATGCCTAATGGTTCATTGCTAGACAGACTCTCTTGCTTGGGGATGGTACT  889

Qy  281  ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn  300
         ||||||||||||||||||||||||||||||||||||||
Db  890  CCACCACTTTCTTGGCACATGAGATGCAAGATTGCTCAGGGTGCAGCTAATGGCATCAAT  949

Qy  301  PheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp  320
         ||||||||||||||||||||||||||||||||||||||
Db  950  TTTCTACATGAAAATCATCATATTCATAGAGATATTAAAAGTGCAAATATCTTACTGGAT  1009

Qy  321  GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla  340
         |||||||||||||||||||||||||||||||||||
Db 1010  GAAGCTTTTACTGCTAAAATATCTGACTTTGGCCTTGCACGGGCTTCTGAGAAGTTTGCC  1069

Qy  341  GlnThrHisValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu  360
         ||||||||||||||||||||||||||||||||||||||||
Db 1070  CAGACAGTCATGACTAGCAGAATTGTGGGAACAACAGCTTATATGGCACCAGAAGCTTTG  1129

Qy  361  ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle  380
         |||||||||||||||||||||||||||||||||||||
Db 1130  CGTGGGGAAATAACACCCAAAATCTGATATTTACAGCTTTGGTGTGGTTTTACTAGAAATA  1189

Qy  381  IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAspIleLys  400
         |||||||||||||||||||||||||||||||||
Db 1190  ATAACTGGACTTCCAGCTGTGGATGAACACCGTGAACCTCAGTTATTGCTAGATATTAAA  1249

Qy  401  GluGluIleGluAspGluGluGlyLysThrIleGluGluAspTyrIleGluAspLysLysMetAsnAsp  420
         |||||||||||||||||||||||||||||||||||||||||
Db 1250  GAAGAAATTGAAGATGAAGAAGGAAAAACAATTGAAGAAGATTATATAGAAGATAAGAAAATGAATGAT  1309

Qy  421  AlaAspSerThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys  440
         |||||||||||||||||||||||||||||||||||||||
Db 1310  GCTGATTCCACTTCAGTTGAAGCTATGTACTCTGTGGCTAGCCAATGTCTGCATGAAAAG  1369

Qy  441  LysAsnLysArgProAspIleLysLysValGlnGlnLeuLeuGluMetThrAlaSer  460
         |||||||||||||||||||||||||||||||||
Db 1370  AAAAATAAGAGGCCCAGACATTAAGAAGGTTCACCAGCGTGCAAGAGGATGACAGCTTCT  1429
```

RESULT 4
AAD10198
ID   AAD10198 standard; cDNA; 1542 BP.
XX
AC   AAD10198;
XX
DT   24-SEP-2001 (first entry)
XX
DE   Mouse interleukin-1 (IL-1) receptor-associated kinase (IRAK)-4 cDNA.
XX
KW   Mouse; interleukin-1 receptor associated kinase-4; IRAK-4; cytostatic;
KW   IL; antibacterial; antiinflammatory; ophthalmological; vasotropic; OPD;
KW   chronic obstructive pulmonary disease; neuroprotective; chronic cough;
KW   adult respiratory disease syndrome; pulmonary fibrosis; asthma; ARDS;
KW   interstitial lung disease; allergic rhinitis; transplant rejection;
KW   autoimmune disease; rheumatoid arthritis; systemic lupus erythematosus;
KW   multiple sclerosis; diabetes; cancer; solid tumour; lymphoma; stroke;
KW   cardiovascular disease; atherosclerosis; neurodegenerative disease;
KW   sepsis; osteoarthritis; rash; osteoporosis; psoriasis; dermatitis;
KW   inflammatory bowel disease; gout; Crohn's disease; ulcerative colitis;
KW   Behcet's syndrome; ankylosing spondylitis; IL-1 receptor/Toll receptor;
KW   sarcoidosis; transgenic animal; ss.
OS   Mus sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1542
FT   /*tag= a
FT   /product= "Mouse IRAK-4"
XX
PN   WO200151641-A1.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US01171.
XX
PR   13-JAN-2000; 2000US-0176395.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Wesche H, Li S;
XX
DR   WPI; 2001-451860/48.
DR   P-PSDB; AAE05399.
XX
PT   Novel human interleukin-1 receptor associated kinase polypeptide,
PT   useful for identifying modulators of the polypeptide for treating gout,
PT   asthma, allergic rhinitis, multiple sclerosis and skin cancer
XX
PS   Claim 26; Fig 4; 89pp; English.
XX
CC   The present sequence is a cDNA encoding mouse interleukin (IL)-1 receptor
CC   associated kinase (IRAK)-4. IRAK associate with activated IL-1, IL-18
CC   and other receptors and act to transduce signals originating from the
CC   activated receptors, ultimately leading to a variety of downstream
CC   effects such as nuclear factor (NF)-kappaB activation. The IRAK-4
CC   inhibitors are useful for treating inflammatory diseases such as
CC   pulmonary diseases and diseases of the airway (e.g., adult respiratory
CC   disease syndrome (ARDS), chronic obstructive pulmonary disease (OPD),
CC   pulmonary fibrosis, interstitial lung disease, asthma, chronic cough
CC   or allergic rhinitis), transplant rejection, autoimmune diseases (e.g.,

CC   rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis
CC   or diabetes), cancer (e.g., solid tumours, skin cancer or lymphoma),
CC   cardiovascular diseases (e.g., stroke and atherosclerosis), diseases
CC   of the central nervous system (e.g., neurodegenerative disease), CD14
CC   mediated sepsis, non-CD14 mediated sepsis, osteoarthritis, osteoporosis,
CC   psoriasis, diseases of the skin (e.g., rash, contact dermatitis, atopic
CC   dermatitis), inflammatory bowel disease (e.g., Crohn's disease and
CC   ulcerative colitis), Behcet's syndrome, ankylosing spondylitis, gout,
CC   sarcoidosis and ophthalmic diseases and conditions. The inhibitors of
CC   IRAK-4 activity or expression are used to inhibit signal transduction
CC   resulting from the activation of an interleukin-1 receptor (IL-1R)/Toll
CC   receptor in a cell. They also inhibit the activation of a transcription
CC   factor that activates NFkappaB in the cell. IRAK-4 is used to create a
CC   nonhuman transgenic animal which is useful for testing the function of
CC   IRAK-4 in vivo, to generate models for the study of inflammatory
CC   disorders and conditions and for the development of potential treatments
CC   for IRAK-4-related inflammatory diseases and conditions. IRAK-4 sequences
CC   are also used in gene therapy and in antisense therapy.
XX
SQ   Sequence 1542 BP; 421 A; 392 C; 423 G; 306 T; 0 other;

US-10-001-254-16 (1-460) x AAD10198 (1-1542)

Alignment Scores:

Pred. No.:   3.96e-196
Score: 2017.00
Percent Similarity: 90.85%
Best Local Similarity: 83.88%
Query Match: 84.53%
DB: 22

Length: 1542
Matches: 385
Conservative: 32
Mismatches: 42
Indels: 0
Gaps: 0

Qy   1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
            ||||||||||::||||||||||||||||||||||||||||||:::|||:||||||||||||
Db   163 ATGAACAAGCCGTTCACACCATCGACATACACGTACGCAACCTTAATGTGGGAGGATCTTAGG 222

Qy   21 LysLeuSerAspPheIleAspProGlyGluGlyTrpLysValLeuAlaValAlaIleLeuLys 40
            |||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db   223 AAGCTGTCCGGATTTATTGATCCCGGAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAA 282

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
            |||||||||||::|||||||||||||||||||||||||:||||||||||||||
Db   283 AAGCCCTCCGGCGACGACAGATACAATCAGTTCCATATAAGGAGATTCGAAGCCTTACTT 342

Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   343 CAGACGGGGAAGAGCCCCACCTGTGAACTGCTGTTTGACTGGGGCACCACGAACTGCACA 402

Qy   81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGlyPheHisAlaProAlaSerLeuLeu 100
            ||||||::||||||||||||||||||||||||||||||||||||||
Db   403 GTTGGGGACCTTGTGGATCTACTGGGTCCAGATTGACGTGTTTCCCCCGCCACTCTCCTG 462

Qy   101 LeuProAspAlaValProLysGlyThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120
            ||||||||||||||||||||||||::|||||||||||:::||||
Db   463 CTGCCCGGATGCCGTTCCCCAAACCGGTCAAAACCTGCCTCCTTAGGAGACGCGCAACAGTG 522

Qy   121 GlnGlnLysGlnMetProPheCysAspGlyAspArgThrLeuMetThrProValGlnAsn 140
            |||||||::|||||||||||||||||||||:::|||||||||||||||
Db   523 GCACAAACACAGGGGCCTTGTCAGGAAAAGACAGGACATCCGTAATGCCTATGCCGAAG 582

Qy   141 LeuGluGlnSerTyrMetProProAspSerSerProGluAsnLysSerLeuGluVal 160
            |||||||||::|||||||||||||||||||||||||||||||:::||||
Db   583 CTAGAACACACGCTGCCAGCCACGGACTCCTCCAAGGCCCAACGAACAAGAGTGTAGAGTCC 642

Qy   161 SerAspThrArgPheHisSerPheTyrGluLysLysAsnValThrAsnAsnPhe 180
            |||||||||||:::|||||||||||||||||||||||||||||||
Db   643 AGCGACACTCGGTTCCACAGCTTCTCGTTCAAGAAGCATCAAACACACTTC 702

Qy   181 AspGluArgProIleSerGlnAlaGlyAsnLysMetGlyGlyGlyPheGlyValVal 200
            ||||:::|||||||||||||||||||||||||||||||||||||||
Db   703 GACGAGCAACCCGCGTCTGCGGGTGGCAACAGCCGGAGGGGATTGGAGTGGTG 762

Qy   201 TyrLysGlyTyrValAsnAsnThrThrValAlaValAlaLysAlaLeuAlaMetValAsp 220
            ||||||||||||||||||||||||||||||||||||||||::::
Db   763 TACAAGGGCTGTGTGAACAACACCACCGTGGCGGTGGTGAAGAAGCTCGGAGCGATGGTTGAA 822

Qy   221 IleThrThrGluGlnLeuLeuLysGlnGlnPheGluGlnLeuIleLysGluValMetAlaLysCys 240
            |||:::|||||||||||||||||::::|||||||||||||||||||||||||||||||
Db   823 ATCAGTACTGAAGGACTAAAGCAACAGTTTGATCAAGAAAATTAAAGTAAGTGGCCAAGTGT 882

Qy   241 GlnHisGlyAsnLeuValGluLeuGlyPheSerSerAspGlyAspAspLeuCysLeu 260
            |||||||||||||||||||||||||||||||||||||||||::::|||||
Db   883 CAGCACGGGAAACCTGGTGGAGCTGGGCTTCTCCAGCGACACCTGTGCTTA 942

Qy   261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgArgLeuSerLysValGlyThr 280
            ||||||||||:::||||||||||||||||||||||||||||||::::|||||
Db   943 GTGTATGTCTACATGCCCAACGGGTCCCTGCTGGACAGGACGTTATCCAAAGTGGGTACA 1002

Qy   281 ProProLeuSerTrpHisMetArgCysLysIleAlaLeuGlnIleAlaLeuAlaAsnGlyLeuAsn 300
            ||||||||||||::|||||||||||||||||||||||||||||||||||::::
Db   1003 CCACCCGCTTCCTTGGCCACAAGGGTGCCAAGGTCGCCCTGGAGGGACACGAGACTGTCCTGG 1062

Qy   301 PheLeuHisGluAsnHisHisIleHisArgAspIleLysAlaAsnIleLeuLeuAsp 320
            ||||||||||::::|||||||||||||||||||||||||||||||||
Db   1063 TTTCTGCATGAAAAACATCACATTCATTAGAGATATTAAAAGTGCAAATATTCTTACTAGAC 1122

Qy   321 ArgGlyThrIleThrProSerArgPheGlyLeuAlaArgAlaSerGluGlyPheAla 340
            ||||||||||||||||||||::::||||||||||||||||||||||::
Db   1123 CGGGGACGAATAACACCAAAATCTGACATTGCAAAAAAATCTGACAGGTTCGGTTTGGAGCTG 1182

Qy   341 GlnThrValMetThrSerArgIleValGlyThrArgTyrMetAlaProGluAlaLeuAsn 360
            |||||||||||||:::||||||||||||||||||||||||||::::|||||
Db   1183 CAGACCGTCATGACCAGCCGAATCGTGGGCACAAGGTACATGGCACCCGAAGCTTTG 1242

Qy   361 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluGlyPheAla 380
            :::||||||||||||||||||||||||||||||||||||::::||||||
Db   1243 GAAGCATTTACTGCCAAAAATATCTGACTTTGGGCTTGCAAGGGCTAGCG 1302

Qy   381 IleThrThrGlyLeuProAlaValAlaGlnHisArgGlnProGlnLeuLeuLeuAspAspIleLys 400
            ||||||||||||||||||||||||::::||||||||||||||||||||||||||||
Db   1303 ATAACCACGGGCTGCCGGCTGTGGCTCAGCATCGACCCTCAACTACTGCTGGATATTAAA 1362

Qy   401 GluGluIleLeuAspGluLeuThrGlnLysThrIleLeuAspTyrIleGluAspIleLysSerMetAsnAsp 420
            |||||||||||||||||||||||||||||:::|||||||||||||||||||
Db   1363 GAAGAGATTGAAGATGAGAAGAGACAGATTAAAGATTACAGGATGAGAAGATAGGGAT 1422

Qy   421 AlaAspSerThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys 440
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1423 GCGGACCCTGCTTCGGTGGAAGCAAATGTACTACTGCGTCTGCAGCCAGTGCATGAGAAG 1482

Qy   441 LysAsnLysArgProAlaProIleLysLysLysValGlnLeuLysGlnLeuMetThrAla 459
            |||||:::||||||||||||||||||||||||||||||||||:::|||
Db   1483 AAAAACAGACGGCCCAGACATTGCAAAGGTTCAACAGCTGACACAAGAGATGTCTGCT 1539

RESULT 5
AAH13798
ID   AAH13798 standard; cDNA; 2213 BP.
XX
AC   AAH13798;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:10742.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; 88.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.

```
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 10742; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the complementary sequences defined in the specification, where the
CC   5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH13742 represent human cDNA sequences; AAB92446 to
CC   AAB95891 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2213 BP; 642 A; 488 C; 516 G; 567 T; 0 other;

     US-10-001-254-16 (1-460) x AAH13798 (1-2213)

Alignment Scores:
Pred. No.:          1.88e-172      Length:        2213
Pred:               1                 
Score:              1787.50         Matches:       2213
Percent Similarity: 87.34%         Conservative:  352
Best Local Similarity: 87.34%      Mismatches:    0
Query Match:        74.92%         Indels:        1
DB:                 22             Gaps:          1
```

```
Qy      1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  751 ATGAACAAGAAACCCATAACACCATCAACACATATGTGCGGTGCCTCAAATGTTGGACTAATTAGG 810

Qy     21 LysLeuSerAspPheIleAspProGlnGlyTrpLysLysLeuAlaValAlaAlaLeuLys  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  811 AAGCTGTCAGATTTTATTGATCCTCAAGGATGGAAGAAGTTAGCTGTGTAGCTATTAAA 870

Qy     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  871 AAACCATCTGGTGATGATAGATAGCAATCAGTTTCACATAAGA 912

Qy     61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  912 ---------------------------------------------------------- 912

Qy     81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeu  100
          -----------------------------------------------------------
Db  912 ---------------------------------------------------------- 912

Qy    101 LeuProAspAla-ValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVa  120
          ||||||||||||-||||||||||||||||||||||||||||||||||||||||||||||
Db      ----------------------------------------------------------

Qy    120 IGlnGlyLysGlnMetProThrCysAspAspGlyArgThrLeuMetThrProValGlnAs  140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  913 ------------TGCTGTTCCCAAAACTGCTAATACACTACCTTCTAAAGAAGCTATAACAGT  963

Qy    140 nLeuGluGlnIleSerTyrMetProProProAspSerSerSerProGluAsnLysSerLeuGluVa  160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  964 TCTTGAACAAAAGCTATAATGCCACCTGACTCCTCCAAGTCCCAGAAAAATAAAAGTTAGAAGT 1023

Qy    160 lSerAspThrArgPheHisSerSerPheTyrGlnLeuValThrAsnAsnPh  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1024 TCATTAAGACTTTCACAGTTTTCATTTTATGAATTGAACAAAGTTCAAAATAACTT 1083

Qy    180 eAspGluArgProIleSerGlnValGlyGlyAsnLysMetGlyGluGlyGlyValVa  200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1084 TAGTGATACACGTTTTCACAGTTTTCATTTTATGAATTGAACAAAGTTCAAAATAACTTI 1143

Qy    200 lTyrLysGlyTyrValAsnAsnThrThrValAlaValAlaAlaIleMetValAs  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1144 TGATGAACGACCCCATTTCTGTTGGTGGTAATAAAAATGGGAGAGGAGGAGTTGTGT 1203

Qy    220 pIleThrThrGlyEluGluLeuLysGlnInPheAspGlnIleLeuGluLysValMetAlaLysCy  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1204 ATATAAAGGCTACGTAAAATAACACAACTGTGGCCAGTGAAGAGCTTGCAGCAATGGTTGA 1263

Qy    240 sGlnHisGluAsnLeuValGlyLeuLeuLeuGlyPheSerSerAspGlyAspProLeuCysLe  260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1264 CATTACTACTGAGGAACTGAAAACAGCAGTTTGATCAAGAAAATAAAGGTGCAAAGTG 1323

Qy    260 uValThrTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyTh  280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1324 TCAACACATGAAAACTTAGTGAACTACTTGGTTTCTCAAGTGATCGGATGACCCTGCTT 1383

Qy    280 rProProLeuSerTrpHisMetArgCysLysLysIleAlaGlnAlaAlaAsnGlyIleAs  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1384 AGTTATATGTTACATGCCTAACGGTTCAATGCCTAGACACAGACTCTTGCTTGGATGGTAC 1443

Qy    300 nPheLeuHisGluAsnHisHisIleHisArgAspSerAlaAsnIleLeuLeuAs  320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1444 TCCACCACTTTCTTGGCACATGAGAATGCTAGAGATGGAAGATTGCTCAGCTGTAATGGCATCAA 1503

Qy    320 pGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAl  340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1504 TTTTCTACATGAAAATCATCATTACATTCCAATTTCATGAGACTATTTAAAAGTCCAAATATCTTTACTCGA 1563

Qy    340 aGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluIleLe  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1564 TGAAGCTTTACTGCTAAAATATCTGACTTTGGCCTTGCACGGGCTTCTGAGAAGTTTGC 1623

Qy    360 uArgGlyGlueThrProLeuSerAspIleTyrSerPheGlyValValLeuLeuGlul  380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1684 GCCTGGAGAAAATAACACCCAAACTGTGATTATTTACAGCTTTGGTGTGGTTTTACTAGAAAT 1743

Qy    380 ellleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuAspIleLy  400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1744 AATAAACTGGACTTCCACGGTGTGGAATGAACACCGTAACCTCAGTTATTGCTAGATATTAA 1803

Qy    400 sGluGlu 402
          |||||||
Db 1804 AGAAGAA 1810
```

```
RESULT 6

ID    AAS76805
XX
ID    AAS76805 standard; cDNA; 1668 BP.
XX
AC    AAS76805;
XX
DT    13-FEB-2002  (first entry)
XX
DE    DNA encoding novel human diagnostic protein #12609.
XX
```

KW Human; chromosome mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG12618.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID No 12609; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1668 BP; 571 A; 289 C; 332 G; 476 T; 0 other;

Alignment Scores:
Pred. No.:

Score:	1.3e-132	Length:	1668
	1396.00	Matches:	288
Percent Similarity:	72.38%	Conservative:	16
Best Local Similarity:	58.57%	Mismatches:	22
Query Match:	58.51%	Indels:	94
DB:	23	Gaps:	5

US-10-001-254-16 (1-460) x AAS76805 (1-1668)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg   20
           ||||||||||||||||||||||||||||
Db   1 ATGAACAAAACCCATAACACATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG   60

Qy   21 LysLeuSerSerAppPheIleAspProGlnGluGlyThrTrpLysLysLeuAlaValAlaLeuLys   40
           |||||||||||||||||||||||||||||||||
Db   61 AAGCTCTCAGATTTTATTGATCCTCAAGAAGGAAGAACAGTTAGCTGTGTAGCTATTAAA   120

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu   60
           |||||||||||||||||||||||||||||||||||::
Db   121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTT   180

Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeuPheGlyThrThrAsnCysThr   80
           ||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAAAACTGGAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGCACCACAAATTGCACA   240

Qy   81 ValGlyAspLeuValAspProLeuLeuIleGlnAsnGluPheAlaProAlaSerLeuLeu   100
           ||||||||||||||||||||||||||||||||||:::
Db   241 GTTGGGTGATCTGTTGGATCCTTTTGATCCAAAAGAATTTTTTGCTCCTGCAGTCTTTTG   300

Qy   101 LeuProAspAlaValProLysThrAlaAlaAsnThrLeuProSerLysGluAlaIleThrVal   120
           ||||||||||||||||||||||||||||||||||||||||||
Db   301 CTCCCAGATGCTGTTCCCAAAACTGCTAATACACTTCCTTAAAGAAGCTATAACAGTT   360

Qy   121 GlnGlnLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn   140
           ||||||||||||||||||||||||||||||||||||||||
Db   361 CAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGGACATTGATGACACCTGTGCAGAAT   420

Qy   141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerSerLeuGlyVal   160
           ||||||||||||||||||||||||||||||||:::
Db   421 CTTGAACAAAGCTAATAGCCAACCTGACTCCTCAAGTCCAGAAAATAAAAGTTTAGGAGTT   480

Qy   161 SerAspThrArgPheHisSerPhePheTyrGluLeuLysAsnValThrAsnAsnPhe   180
           |||||||||||||||||||||||||||||
Db   481 AGTGATACACGTTTTCACAGTTTTTTCATTTTATGAATTGAAGAATGTCACAAATAACTTT   540

Qy   181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyGlyLysGlyPheGlyValVal   200
           ||||||||||||||||||||||||||||||||||||::
Db   541 GATGAACGAACACCCATTTCTGTTGGTGGTAATAAAATGGGAGGAAAGGGATTTGGAGTTGTA   600

Qy   201 TyrArgGlyTyrValAsnAsnThrThrValAlaValAlaValValAlaLaMetValAsp   220
           ||||||||||||||||||||||||||||||||:::
Db   601 TATAGAGGCTACGTAAATAACACAACTGTGGCAGTGAAGAAGCCTTGCAGCAGTAAGTTAT   660

Qy   221 IleThrThrGlnGluLeuLysGlnPheAspGluIleLeuLysValMetAlaLysCys   240
           ||| |||
Db   661 ATTTTCAGGAATAAAAAGAAA   GAGTTGCTT   681

Qy   241 GlnHisGluAsnLeuValGlyLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu   260
           ||:|||||||||||
Db   682 GAGTTGCTT   690

Qy   261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr   280
           |||||||
Db   691 CATAGTGTG   690

Qy   281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn   300

Db   690   690

Qy   301 PheLeuHisGluAsnHisHisIleHisArgAspLysSerAlaAsnIleLeuLeuAsp   320
           ||:::
Db   690 CCCTATATAGTGTG   729

Qy   321 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla   340
           |||||||||||||||||||||||||||||||||:::
Db   730 GAAGCTTTTACTGCTAAAATAAATCTGATTTTTGGCCTTGCACGGGCTTCTGAGAAGTTTGCC   789

Qy   341 GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu   360
           |||||||||||||||||||||||||||||||
Db   790 CAGACAGTCATGACTTCACGAATTGTGGGAACAACAGCTTATATGGCACCAGAAGCTTTG   849

Qy   361 ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle   380
           |||||||||||||||||||||||||||||
Db   850 CGTGGAGAAATAACACCCAAAATCTGATATTTGACAGTTT-------   888

Qy   381 IleThrGlyLeuProAlaValAspGluHisAsnLeuAsnIleLys   400
           :::
Db   889 ------------------------GACCATCAA-----------GCAACTATCGTTGATGGAACT   918

Qy   401 GluGluIleGluAspGluLysThrIleGluAspLysSerLysLysMetAsnAsp   420
           |||:::::::
Db   919 TCCAAAATGTGAAGAAAAAGCTGAGTGTTTTAAGCTCCTTTTGGATGAAAAAAACTCATGAG   978

RESULT 7

AAA09319

ID   AAA09319 standard; DNA; 833 BP.
XX
AC   AAA09319;
XX
DT   10-AUG-2000 (first entry)
XX
DE   Human cancer associated antigen precursor DNA, clone NY-REN-64.
XX
KW   renal cancer; cancer associated antigen precursor; diagnosis;
KW   cytostatic; 88.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   50..670
FT   /*tag= a
XX
PN   WO20020587-A2.
XX
PD   13-APR-2000.
XX
PF   04-OCT-1999; 99WO-US22873.
XX
PR   05-OCT-1998; 98US-0166300.
PR   05-OCT-1998; 98US-0166350.
XX
PA   (LIJDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y, Gout I, Tureci O, Sahin U, Pfreundschuh M, Scanlan MJ;
PI   Stockert E, Chen Y, Old LJ, Jager E, Knuth A;
XX
DR   WPI; 2000-303774/26.
DR   P-PSDB; AAY92347.
XX
PT   Preventing, diagnosing and/or treating disorders associated with
PT   abnormal expression of human cancer associated antigens
XX
PS   Claim 57; Page 85; 121pp; English.
XX
CC   AAA09310-20 are novel genes isolated by SEREX screening from a renal
CC   cancer cell line 1973/10.4. The genes encode cancer associated antigen
CC   precursors. These gene products are useful in methods for preventing,
CC   diagnosing and/or treating disorders, especially cancer, associated with
CC   abnormal expression of human cancer associated antigens. The method
CC   comprises contacting a sample from a subject with an agent that
CC   specifically binds to the nucleic acid molecule or expression product
CC   (or fragment) complexed with a human leukocyte antigen (HLA) molecule
CC   and determining the interaction between the agent and the nucleic acid
CC   molecule or the expression product as a determination of the disorder.
XX
SQ   Sequence 833 BP; 273 A; 155 C; 179 G; 226 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 7.78e-110
Best Local Similarity: 1169.00
Percent Similarity: 93.21%
Query Match: 48.99%
DB: 21

US-10-001-254-16 (1-460) x AAA09319 (1-833)

```
Qy       1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 60

Qy      21 LysLeuSerAspPheIleAspProGlnGlyGlyTrpLysLysLeuAlaValAlaIleLys 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50 ATGAACAAACCCATAACACCATCAACATATGTCCGCTGCCTCAATGTTGGACTAATTAGG 109

Qy      21 LysLeuSerAspPheIleAspProGlnGlyGlyTrpLysLysLeuAlaValAlaIleLys 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     110 AAGCTGTCAGATTTTATTGATCCTCAAGGAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 169

Qy      41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     170 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGATTTGAAGCATTACTT 229

Qy      61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrArgThr 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     230 CAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCACCACAAGTACA 289

Qy      81 ValGlyAspLeuValAspLeuLeuLeuIleGluIleAsnGluPhePheAlaProAlaSerLeuLeu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     290 GTTGGTGATCTTGTTGGATCTTTGATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTG 349

Qy     101 LeuProAspProAlaValProLysThrAlaAsnThrLeuProSerLeuGluAlaIleThrVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     350 CTCCCAGATCCAGCTGTTCCCAAAACTGCTAATACACCTTCTAAAGAAGCTATAACAGTT 409

Qy     121 GlnGlnLysGlnMetProPheCysTyrAspLysAspArgThrLeuMetThrProValGlnAsn 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     410 CAGCAAAAACAGATGCCTTTCTGCTACGACAAAGACAGGACATTGATGACACCTGTCCAGAAT 469

Qy     141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     470 CTTGAACAAAGCTATATGCCACCAGATTCTTCTTCTCCAGAAAATAAAAGTTTAGAAGTT 529

Qy     161 SerAspThrArgPheHisSerPheSerPheTyrGluLeuLeuArgAsnValThrAsnAsnPhe 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     530 AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGCTAAGAAATGTCACAAATAACTTT 589

Qy     181 AspGluArgProIleSerValGlyGlyLysAsnLysMetGlyGlyLysGlyGlyValVal 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     590 GATGAACGACCCATTTCTGTTGGTGGTAATAAAATGGGAGGGAAGGGAGGTTGTTGTA 649

Qy     201 Tyr-LysGlyTyrValValAsnAsnThrThrValAlaVal-LysLysLeuAlaAlaMetValAla 220
           |||  |||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db     650 TATAAAGGGCTACGTAGTAAATAACACAACTGTGGCAGTGAAAAAGCTTGCAGCAATGGTTG 709

Qy     220 SerIleThrThrGlu-GluGlnPro-GlnGlnGlu-IleLeuLysValMetAla 238
           |||||||||||||| ||||||||  ||||||||  ||||||||||||||||||
Db     710 ACATTACTACTGAAGGAACTGAAACCAGTTGATCCAAGAAAATAAAAGTAATGGC 769

Qy     239 LysCysGlnHis------GluAsnLeuVal---GluLeuLeuGlyPheSerSer---Asp 254
           ||| :||:||      |||||||| |||   ||||||||||||||||  ||| |||
Db     770 AAAAGTGGTCAACCATGGAAAAAACTTTAGTAGGAACCTTACTTGGGTTTTCTCAAGTGGAT 829

Qy     255 Gly 255
           |||
Db     830 GGG 832
```

RESULT 8

AAS27214/c

ID   AAS27214 standard; cDNA; 1493 BP.
XX
AC   AAS27214;
XX
DT   07-NOV-2001 (first entry)
XX
DE   cDNA encoding novel signal transduction pathway protein, Seq ID 249.
XX
KW   Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW   antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW   immune system disorder; rheumatoid arthritis; inflammatory condition;
KW   organ transplant rejection; infection; hepatitis C; blood disorder;
KW   sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW   cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW   reproductive system; gastrointestinal; liver disorder; AIDS; 88;
KW   acquired immune deficiency syndrome.
XX
OS   Homo sapiens.
XX
PN   WO20015473733-A1.
XX
PD   02-AUG-2001.

```
XX
PF  17-JAN-2001;  2001WO-US01312.
XX
PR  31-JAN-2000;  2000US-0179065.
PR  04-FEB-2000;  2000US-0180628.
PR  24-FEB-2000;  2000US-0184664.
PR  02-MAR-2000;  2000US-0186350.
PR  16-MAR-2000;  2000US-0189874.
PR  17-MAR-2000;  2000US-0190076.
PR  18-APR-2000;  2000US-0198123.
PR  19-MAY-2000;  2000US-0205515.
PR  28-JUN-2000;  2000US-0209467.
PR  07-JUN-2000;  2000US-0214886.
PR  30-JUN-2000;  2000US-0215135.
PR  07-JUL-2000;  2000US-0216647.
PR  07-JUL-2000;  2000US-0216880.
PR  11-JUL-2000;  2000US-0217487.
PR  11-JUL-2000;  2000US-0217496.
PR  14-JUL-2000;  2000US-0218290.
PR  14-JUL-2000;  2000US-0220963.
PR  26-JUL-2000;  2000US-0220964.
PR  26-JUL-2000;  2000US-0224518.
PR  14-AUG-2000;  2000US-0224519.
PR  14-AUG-2000;  2000US-0225213.
PR  14-AUG-2000;  2000US-0225214.
PR  14-AUG-2000;  2000US-0225266.
PR  14-AUG-2000;  2000US-0225267.
PR  14-AUG-2000;  2000US-0225268.
PR  14-AUG-2000;  2000US-0225270.
PR  14-AUG-2000;  2000US-0225447.
PR  14-AUG-2000;  2000US-0225757.
PR  14-AUG-2000;  2000US-0225758.
PR  18-AUG-2000;  2000US-0225759.
PR  22-AUG-2000;  2000US-0226279.
PR  22-AUG-2000;  2000US-0226681.
PR  22-AUG-2000;  2000US-0226868.
PR  23-AUG-2000;  2000US-0227009.
PR  30-AUG-2000;  2000US-0227182.
PR  01-SEP-2000;  2000US-0228924.
PR  01-SEP-2000;  2000US-0229287.
PR  01-SEP-2000;  2000US-0229343.
PR  01-SEP-2000;  2000US-0229344.
PR  05-SEP-2000;  2000US-0229345.
PR  05-SEP-2000;  2000US-0229509.
PR  05-SEP-2000;  2000US-0229513.
PR  06-SEP-2000;  2000US-0230437.
PR  06-SEP-2000;  2000US-0230438.
PR  08-SEP-2000;  2000US-0231242.
PR  08-SEP-2000;  2000US-0231243.
PR  08-SEP-2000;  2000US-0231244.
PR  08-SEP-2000;  2000US-0231413.
PR  08-SEP-2000;  2000US-0231414.
PR  08-SEP-2000;  2000US-0231968.
PR  12-SEP-2000;  2000US-0232081.
PR  14-SEP-2000;  2000US-0232080.
PR  14-SEP-2000;  2000US-0232297.
PR  14-SEP-2000;  2000US-0232298.
PR  14-SEP-2000;  2000US-0232299.
PR  14-SEP-2000;  2000US-0232400.
PR  14-SEP-2000;  2000US-0232401.
PR  14-SEP-2000;  2000US-0233063.
PR  14-SEP-2000;  2000US-0233064.
PR  14-SEP-2000;  2000US-0233065.
PR  21-SEP-2000;  2000US-0234223.
PR  21-SEP-2000;  2000US-0234274.
PR  25-SEP-2000;  2000US-0234997.
PR  25-SEP-2000;  2000US-0234998.
PR  26-SEP-2000;  2000US-0235484.
PR  27-SEP-2000;  2000US-0235834.
PR  27-SEP-2000;  2000US-0235836.
PR  29-SEP-2000;  2000US-0236327.
PR  29-SEP-2000;  2000US-0236367.
PR  29-SEP-2000;  2000US-0236368.

PR  29-SEP-2000;  2000US-0236369.
PR  29-SEP-2000;  2000US-0236370.
PR  02-OCT-2000;  2000US-0236802.
PR  02-OCT-2000;  2000US-0237037.
PR  02-OCT-2000;  2000US-0237038.
PR  02-OCT-2000;  2000US-0237039.
PR  02-OCT-2000;  2000US-0237040.
PR  13-OCT-2000;  2000US-0239935.
PR  13-OCT-2000;  2000US-0239937.
PR  20-OCT-2000;  2000US-0240960.
PR  20-OCT-2000;  2000US-0241221.
PR  20-OCT-2000;  2000US-0241785.
PR  20-OCT-2000;  2000US-0241786.
PR  20-OCT-2000;  2000US-0241787.
PR  20-OCT-2000;  2000US-0241808.
PR  20-OCT-2000;  2000US-0241809.
PR  20-OCT-2000;  2000US-0241826.
PR  01-NOV-2000;  2000US-0244617.
PR  08-NOV-2000;  2000US-0244674.
PR  08-NOV-2000;  2000US-0244675.
PR  08-NOV-2000;  2000US-0244476.
PR  08-NOV-2000;  2000US-0246475.
PR  08-NOV-2000;  2000US-0246476.
PR  08-NOV-2000;  2000US-0246477.
PR  08-NOV-2000;  2000US-0246478.
PR  08-NOV-2000;  2000US-0246523.
PR  08-NOV-2000;  2000US-0246524.
PR  08-NOV-2000;  2000US-0246525.
PR  08-NOV-2000;  2000US-0246526.
PR  08-NOV-2000;  2000US-0246527.
PR  08-NOV-2000;  2000US-0246528.
PR  08-NOV-2000;  2000US-0246532.
PR  08-NOV-2000;  2000US-0246609.
PR  08-NOV-2000;  2000US-0246610.
PR  08-NOV-2000;  2000US-0246611.
PR  08-NOV-2000;  2000US-0246613.
PR  17-NOV-2000;  2000US-0249207.
PR  17-NOV-2000;  2000US-0249208.
PR  17-NOV-2000;  2000US-0249209.
PR  17-NOV-2000;  2000US-0249210.
PR  17-NOV-2000;  2000US-0249211.
PR  17-NOV-2000;  2000US-0249212.
PR  17-NOV-2000;  2000US-0249213.
PR  17-NOV-2000;  2000US-0249214.
PR  17-NOV-2000;  2000US-0249215.
PR  17-NOV-2000;  2000US-0249216.
PR  17-NOV-2000;  2000US-0249217.
PR  17-NOV-2000;  2000US-0249218.
PR  17-NOV-2000;  2000US-0249244.
PR  17-NOV-2000;  2000US-0249245.
PR  17-NOV-2000;  2000US-0249264.
PR  17-NOV-2000;  2000US-0249265.
PR  17-NOV-2000;  2000US-0249297.
PR  17-NOV-2000;  2000US-0249299.
PR  17-NOV-2000;  2000US-0249300.
PR  01-DEC-2000;  2000US-0250160.
PR  01-DEC-2000;  2000US-0250391.
PR  05-DEC-2000;  2000US-0251030.
PR  05-DEC-2000;  2000US-0251479.
PR  05-DEC-2000;  2000US-0256719.
PR  06-DEC-2000;  2000US-0251856.
PR  08-DEC-2000;  2000US-0251868.
PR  08-DEC-2000;  2000US-0251869.
PR  08-DEC-2000;  2000US-0251989.
PR  08-DEC-2000;  2000US-0251990.
PR  11-DEC-2000;  2000US-0254097.
PR  05-JAN-2001;  2001US-0259678.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Rosen CA, Barash SC, Ruben SM;
XX
DR  WPI; 2001-465460/50.
DR  P-PSDB; AAU17297.
```

XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
PS Claim 1; SEQ ID No 249; 880pp; English.
XX

CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (II), (III) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g arrhythmia) respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX

Alignment Scores:
Pred. No.:
Score: 972.0
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 40.74%

2.4e-89

Length: 1493
Matches: 190
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-001-254-16 (1-460) x AAS27214 (1-1493)
DB:

Qy 271 LeuAspArgLeuSerCysLeuAspGlyThrProLeuSerTrpHisMetArgCysLys 290
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1491 CTAGACAGACTCTCTTGCTTGGATAGGGTACTCCACCACTTTCTTGGCACATGAGATGCAAG 1432

Qy 291 IleAlaGlnGlyAlaAlaAsnGlyIleAsnPheLeuHisGluAsnHisHisIleHisArg 310
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1431 ATTGCTCAGGGTGCAGCTAATGGCATCAATTTCTACATGAAAATCATCATAATTCATAGA 1372

Qy 311 AspIleLysSerAlaAsnIleLeuLeuAspGluAlaPheThrAlaLysSerAspPhe 330
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1371 GATATTAAAAGTGCAAATATCTTACTCGATGAAGCTTTTACTGCTAAAATATCTGACTTT 1312

Qy 331 GlyLeuAlaArgAlaSerGlyLysPheAlaGlnThrValMetThrSerArgIleValGly 350
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1311 GGCCTTGCACGGCTTCTGAGAAGTTTGCCCAGACAGTCATGACTAGCAGAATTGTGGGA 1252

Qy 351 ThrThrAlaTyrMetAlaProGluAlaLeuAsnGlyLysLeuThrGluLysSerAspIle 370
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1251 ACAACAGCTTATATGGCACCAGAAGCTTTGCCTGGAAAATAACACCCAAAATCTGATATT 1192

Qy 371 TyrSerPheGlyValValLeuLeuGluLeuIleThrGlyLeuProAlaValAlaAspHis 390
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1191 TACAGCTTTGGTGTGGTTTTACTAGAAAATAATAACTGGACTTCCAGCTGCTGATGAACAC 1132

Qy 391 ArgGluProGlnLeuLeuLeuAspIleIleLysGluLysGluLysThrIle 410
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1131 CGTGAACCTCAGTTATTGCTAGATATTAAAGAGAAAAAGATGAAACAAATT 1072

Qy 411 GluAspTyrIleAspThrLysLysMetAsnAspAlaAspSerThrSerValGluAlaMetTyr 430
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1071 GAAGATTATATTGATACTAAAAAGATGAATGACGCTGATTCCACTTCAGTTGAAGCTATGTAC 1012

Qy 431 SerValAlaSerGlnCysLeuHisGluLysGlyLysAsnLeuLysArgProAspIleLeuLysGlySerVal 450
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1011 TCTGTTGCTAGTCAATGTCTTCATGAAAAGGGAAAAAATAAGAGACCAGACAGATTAAGAAGGTT 952

Qy 451 GlnGlnLeuLeuGluGlnGluMetThrAlaSer 460
        |||||||||||||||||||||||||||||
Db 951 CAACAGCTGCTGCAAGAGAGATGACAGCTTCT 922

RESULT 9
ABK44003/c
ID   ABK44003 standard; cDNA; 1493 BP.
XX
AC   ABK44003;
XX
DT   05-JUN-2002 (first entry)
XX
DE   DNA encoding novel central nervous system protein #583.
XX
KW   Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW   cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW   acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW   adenocarcinoma; reproductive system disorder; testicular feminisation;
KW   endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW   respiratory disorder; renal disorder; kidney failure; blood disorder;
KW   myocardial infarction; wound healing; cell proliferation; skin aging;
KW   food additive; food preservative; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155318-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01332.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226668.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0227102.
PR   23-AUG-2000; 2000US-0227182.
PR   30-AUG-2000; 2000US-0228924.

```
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0211413.
PR   08-SEP-2000; 2000US-0211414.
PR   08-SEP-2000; 2000US-0230080.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232081.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.

PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.

PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251479.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-581633/65.
DR   P-PSDB; AAU87673.
XX
PT   New isolated nucleic acid encoding a protein for diagnosing,
PT   preventing, treating or ameliorating medical conditions and used as
PT   food additives or preservatives -
XX
PS   Claim 1; SEQ ID No 593; 837pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule (I) encoding a
CC   novel central nervous system protein. (I) and polypeptides (III) encoded
CC   by (I), are used to treat a medical conditions and in diagnosis of a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC   amylotrophic lateral sclerosis, infections caused by bacteria, viruses
CC   e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC   e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC   adenocarcinomas and irritable bowel syndrome, reproductive system
CC   disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC   and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC   leukaemia, disorders involving neovascularisation e.g. malignancies,
CC   respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC   acute kidney failure and blood related disorders e.g. myocardial
CC   infarction. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities, fat content, lipid, protein,
```

Alignment Scores:
Pred. No.:

Pred. No.:	2.4e-89	Length:	1493
Score:	972.00	Matches:	190
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.74%	Indels:	0
DB:	23	Gaps:	0

```
US-10-001-254-16 (1-460) x ABK44003 (1-1493)

Qy  271 LeuAspArgLeuSerCysLeuAspGlyThrProProLeuSerTrpPhisMetArgCysLys 290
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1491 CTAGACAGACTCCCTTGCTTGGATGTACTCCACCACTTTCTTGGCACATGAGATGCAAG 1432

Qy  291 IleAlaGlyIleAlaAlaAsnGlyIleAsnPheLeuHisGluAsnHisHisIleHisArg 310
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1431 ATTGCTCAGGGTGCAGCTAATGGCATCAATTTCTTACATGAAAATCATCATTATTCATAGA 1372

Qy  311 AspIleLysSerAlaAsnIleLeuLeuAspGluAlaPheThrAlaLysIleSerAspPhe 330
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1371 GATATTAAAAGTGCAAATATCTTACTGGATGAAGCTTTTACTGCTAAAATATCTGACTTT 1312

Qy  331 GlyLeuAlaArgAlaSerGlyLysPheAlaGlnThrValMetThrSerArgIleValGly 350
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1311 GGCCTTGCACGGCGTTCGAGAAGTTTGCCCAGACAGTACATGACTAGCAGAATTGTGGGA 1252

Qy  351 ThrThrAlaTyrMetAlaProGluAlaLeuArgGlyGluIleThrProLysSerAspIle 370
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1251 ACAACAGCTTATATGGCACCAGAAGCTTTGCGTGGAGAAATAACACCCAAAATCTGATATT 1192

Qy  371 TyrSerPheGlyValValLeuLeuGluLeuIleThrGlyLeuProAlaValAspGluHis 390
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1191 TACAGCTTTGGTGTGGTTTTACTAGAAAATAATAACTGGACTTCCAGCTGTGGATGAACAC 1132

Qy  391 ArgGluProGlnLeuLeuAspAspIleLysGluGluGluGluLysLysThrIle 410
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1131 CGTGAACCTCAGTTATTGCTAGATATTAAAGAAGAAGAAGAAGAAAAGAAGACAATT 1072

Qy  411 GluAspTyrIleLeuAspLysLysMetAsnAspAlaAspSerThrSerValGluAlaMetTyr 430
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1071 GAAGATTATATTCTAGATAAGAAAATGAATGCTGATTCCACTTCAGTTGAAGCTATGTAC 1012

Qy  431 SerValAlaSerGlnCysLeuHisGlyLeuLysAsnLysArgProAspIleLysLysVal 450
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1011 TCTGTTGCTAGTCAAATGTCTCATGGAAAGAAAAATAAGAGACCAGACATTAAGAAGGTT 952

Qy  451 GlnGlnLeuLeuGlnGluMetThrAlaSer 460
            ||||||||||||||||||||||||||||||
Db  951 CAACAGCTGCTGCAAGAGATGACAGCTTCT 922
```

```
RESULT 10
ID   ABK43727 standard; cDNA; 1618 BP.
XX
AC   ABK43727;
XX
DT   05-JUN-2002  (first entry)
XX
DE   DNA encoding novel central nervous system protein #307.
XX
KW   Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW   cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW   nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW   acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW   adenocarcinoma; reproductive system disorder; testicular feminisation;
KW   endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW   respiratory disorder; renal disorder; kidney failure; blood disorder;
KW   myocardial infarction; wound healing; cell proliferation; skin aging;
KW   food additive; food preservative; gene therapy; gene; ss.
OS   Homo sapiens.
XX
PN   WO200155318-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01332.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0255757.
PR   14-AUG-2000; 2000US-0255758.
PR   14-AUG-2000; 2000US-0255759.
PR   18-AUG-2000; 2000US-0226279.
PR   18-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   12-SEP-2000; 2000US-0232080.
PR   12-SEP-2000; 2000US-0232081.
PR   14-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   14-SEP-2000; 2000US-0234222.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234998.
PR   25-SEP-2000; 2000US-0235484.
PR   26-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
```

XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
DR P-PSDB; AAU87397.
DR WPI; 2001-581633/65.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PR 05-JAN-2001; 2001US-0259678.
PR 11-DEC-2000; 2000US-0254097.
PR 08-DEC-2000; 2000US-0251990.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251856.
PR 06-DEC-2000; 2000US-0251479.
PR 05-DEC-2000; 2000US-0256719.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251030.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0250160.
PR 17-NOV-2000; 2000US-0249300.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249210.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246474.
PR 01-NOV-2000; 2000US-0244617.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0240960.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0239935.
PR 02-OCT-2000; 2000US-0237040.
PR 02-OCT-2000; 2000US-0237039.

Claim 1: SEQ ID No 317; 837pp; English.
PS
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (II) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amylotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:
Pred. No.:
Score: 972.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 40.74%

US-10-001-254-16 (1-460) x ABK43727 (1-1618)

Length: 1618
Matches: 190
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

```
Qy  271 LeuAspArgLeuSerCysLeuAspGlyThrProLeuSerTrpHisMetArgCysLys 290
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3 CTAGACAGACTCTCTTGCTTGGATGGTACTCCACCACTTTCTTGGCACATGAGATGCCAAG 62

Qy  291 IleAlaGlnGlyValAlaAlaAsnGlyIleAsnPheLeuHisGluAsnHisHisIleHisArg 310
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 ATTGCTCAGGGTGCAGCTAATGGCATCAATTTTCTACATGAAAATCATCATATTCATAGA 122

Qy  311 AspIleLysSerAlaAsnIleIleLeuLeuAspGluAlaPheThrAlaIleSerAspPhe 330
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123 GATATTAAAAGTGCAAATATCATTTTACTCGATGAAGCTTTTACTGCTAAAAATTCTGACTTT 182

Qy  331 GlyLeuAlaArgAlaSerGluLysPheAlaGlnThrValMetThrSerArgIleValGly 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183 GGCCTTGCACGGGCTTCTGAGAAGTTTGCCCAGACAGTCATGACTAGCAGAATTGTGGGA 242

Qy  351 ThrThrAlaTyrMetAlaProGluAlaLeuArgGlyGlyIleThrProLysSerAspIle 370
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  243 ACAACAGCTTATATGGCACCAGAAGCTTTGCGTGGAGAAATAACACCCAAATCTGATATT 302

Qy  371 TyrSerPheGlyValValLeuLeuGluIleIleThrGlyLeuProAlaValAspGluHis 390
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  303 TACAGCTTTGGTGTGGTTTTACTAGAAATAATAACTGGACTTCCAGCTGTGGATGAACAC 362

Qy  391 ArgGluProGlnLeuLeuLeuAspIleLysGluGluIleGluAspGluGluLysThrIle 410
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  363 CGTGAACCTCAGTTATTGCTAGATATTAAAGAAGAAATTGAAGATGAAGAAAAGACAATT 422

Qy  411 GluAspTyrIleAspLysLysMetAsnAspAlaSerThrSerValGluAlaMetTyr 430
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  423 GAAGATTATATTGATAAAAAGATGAATGACGCTAGCACTTCAGTTGAAGCTATGTAC 482

Qy  431 SerValAlaSerGlnCysLeuHisGluLysGlyLysSerArgProAspIleLysLysVal 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  483 TCTGTTGCTAGTCAATGCTCGCATGAAAAGGGAAAAAGTAGACCAGACATTAAGAAGGTT 542

Qy  451 GlnGlnLeuLeuGlnGluMetThrAlaSer 460
```

```
Db          RESULT 11

ID          543 CAACAGCTGCTGCAAGAGATGACAGCTTCT 572
            |||||||||||||||||||||||||||||||
            |||||||||||||||||||||||||||||||

ID          ABL90822
AC          ABL90822;
XX
ID          ABL90822 standard; cDNA; 1181 BP.
XX
AC          ABL90822;
XX
DT          24-MAY-2002 (first entry)
XX
DE          Human polynucleotide SEQ ID NO 1384.
XX
KW          Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW          antiallergic; immunosuppressive; hepatotropic; antidiabetic; antiulcer;
KW          vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW          cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW          neurological disease; infection; human; secreted protein; gene; ss.
XX
OS          Homo sapiens.
XX
PN          WO200190304-A2.
XX
PD          29-NOV-2001.
XX
PF          18-MAY-2001; 2001WO-US16450.
XX
PR          19-MAY-2000; 2000US-205515P.
XX
PA          (HUMA-) HUMAN GENOME SCI INC.
XX
PI          Birse CE,   Rosen CA;
XX
DR          WPI; 2002-122018/16.
DR          P-PSDB; ABB90413.
XX
PT          Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT          prevention of neural, immune system, muscular, reproductive,
PT          gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT          disorders -
XX
PS          Claim 4; SEQ ID NO 1384; 208lpp + Sequence Listing; English.
XX
CC          The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC          (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC          medical conditions e.g. by protein or gene therapy. The genes are
CC          isolated from a range of human tissues disclosed in the specification.
CC          The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC          in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC          and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC          marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC          (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC          haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC          disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC          colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC          (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC          .epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC          and parasitic infections.
CC          Note: The sequence data for this patent did not form part of the
CC          printed specification, but was obtained in electronic format directly
CC          from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ          Sequence 1181 BP; 396 A; 222 C; 212 G; 348 T; 3 other;

Alignment Scores:
Pred. No.:
Score:                     3.68e-88
Pred. No.:                 959.00               Length:            1181
Percent Similarity:        98.95%               Matches:           188
Best Local Similarity:     98.95%               Conservative:      0
Query Match:               40.19%               Mismatches:        2
DB:                        24                   Indels:            0
                                                Gaps:              0

US-10-001-254-16 (1-460) x ABL90822 (1-1181)
```

```
Qy    271 LeuAspArgLeuSerCysLeuAspGlyThrProProLeuSerTrpHisMetArgCysLys 290
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 CTAGACAGACTCTCTTGCCTTGGATGGGTACTCCACCACTTTCTTGGCACATGAGATGCAAG 62

Qy    291 IleAlaGlnGlyAlaAlaAsnGlyIleAsnPheLeuHisGluAsnHisHisIleHisArg 310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 ATTGCTCAGGGTGCAGCTAATGGCATCAATTTTCTACATGAAAATCATCATATTCATAGA 122

Qy    311 AspIleLysSerAlaAsnIleLeuLeuAspGluAlaPheThrAlaLysIleSerAspPhe 330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 GATATTAAAAGTGCAAATATCTTACTGGATGAAGCTTTTACTGCTAAAATATCTGACTTT 182

Qy    331 GlyLeuAlaArgAlaSerGlyLysProPheAlaGlnThrValMetThrSerArgIleValGly 350
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 GGCCTTGCACGGGCTTCTGGAAAGCCTTTTGCCCAGACAGTGATGACTAGCAGAATTGTGGGA 242

Qy    351 ThrThrAlaTyrMetAlaProGluValLeuArgGlyGlyIleThrProLysSerAspIle 370
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 ACAACAGCTTATATGGCACCAGAAGTCTTGCGTGGAGGAATAACCCCAAAATCTGATATT 302

Qy    371 TyrSerPheGlyValValLeuLeuGluIleIleThrGlyLeuProAlaValAspGluHis 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 TACAGCTTTGGTGTGGTTTTACTGGAAATAATAACTGGACTTCCAGCTGTGGATGAACAC 362

Qy    391 ArgGluProGlnIleLeuLeuLeuAspAspIleLysGluGluAsnGluGluGlyLysThrIle 410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 CGTGAACCTCAGTTATTGCTGTTAGATGATATTAAAGAAGAGAATGAAGAAGGAAAGACAATT 422

Qy    411 GluAspTyrIleAspLysLysSerMetAsnAspAlaAspSerThrSerValGluAlaMetTyr 430
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 GAAGATTATATTGATAAAAAGAGTATGAATGATGCTGATAGTACTAGCGTTGAAGCTATGTAC 482

Qy    431 SerValAlaSerGlnCysLeuHisGluLysLysAsnLysArgProAspProIleLysLysLysVal 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 TCTGTTGCTAGTCAATGTCTGCATGAAAAGAAAAATAAGAGACCAGACATTAAGAAAGTT 542

Qy    451 GlnGlnLeuLeuGlnGluMetThrAlaSer 460
          ||||||||||||||||||||||||||||||
Db    543 CAACAGCTGCTGCAAGAGATGACAGCTTCT 572
```

```
Db          RESULT 12

ID          AAF22618
AC          AAF22618;
XX
ID          AAF22618 standard; cDNA; 539 BP.
XX
AC          AAF22618;
XX
DT          26-MAR-2001 (first entry)
XX
DE          Human breast cancer associated antigen nucleotide sequence SEQ ID NO:197.
XX
KW          Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW          cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS          Homo sapiens.
XX
PN          WO200073801-A2.
XX
PD          07-DEC-2000.
XX
PF          26-MAY-2000; 2000WO-US14749.
XX
PR          28-MAY-1999;   99US-0136526.
PR          10-SEP-1999;   99US-0153454.
XX
PA          (LUDW-) LUDWIG INST CANCER RES.
XX
PI          Obata Y;
XX
DR          WPI; 2001-025274/03.
XX
PT          Nucleic acids encoding breast, gastric and prostate cancer associated
PT          antigen precursors, useful for diagnosing and treating a condition
```

PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Claim 50; Page 338; 799pp; English.
XX
CC   AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC   represent nucleotide sequences encoding human breast, gastric and
CC   prostate cancer associated antigen precursors (CAAP) respectively.
CC   AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC   represent human breast, gastric and prostate CAAP protein sequence
CC   respectively. CAAPs have cytostatic activity and can be used in the
CC   production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC   acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC   condition characterised by expression of an abnormal amount of a protein,
CC   e.g. cancer.
XX
SQ   Sequence 539 BP; 168 A; 94 C; 111 G; 163 T; 3 other;

US-10-001-254-16 (1-460) x AAF22618 (1-539)

Alignment Scores:
Pred. No.:
Score: 7.17e-81
                          883.00   Length: 539
                                          Matches: 176
Percent Similarity: 97.24%   Conservative: 0
Best Local Similarity: 97.24%   Mismatches: 3
Query Match: 37.01%   Indels: 2
DB: 22   Gaps: 0

Qy   258 LeuCysLeuValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeu 277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CTCTGCTTAGTATATATGTTTACAATGCCTAATGGTTCAATTGCTTAGACAGACTCTTGCTTG 60

Qy   278 AspGlyThrProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsn 297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GATGGTACTCCACCACTTTCTTGGCACATGAGATGCCAAGATTGCTCAGGGTGCCAGCTAAT 120

Qy   298 GlyIleAsnPheLeuHisGluAsnHisHisIleHisArgAspIleTyrSerAlaAsnIle 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGCATCAATTTTCTACATGAAAATCATCATTATTCATAGAGATATTAAAAGTGCAAATATC 180

Qy   318 LeuLeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGlu 337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTACTGGATGAAGCTTTTACTGCTAAAATATCTGACTTTGGCCTTGCACGGGCTTCTGAG 240

Qy   338 LysPheAlaGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaPro 357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AAGTTTGCCCAGACAGTCATGACTAGCAGAATTGTGGGAACACAGCTTATATGGCACCA 300

Qy   358 GluAlaLeuArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeu 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GAAGCTTTGCGTGGTGAGAAAATAACACCCAAAATCTGATATTTACAGCTTTGGTGGTTTTA 360

Qy   378 LeuGluIleLeuThrGlyLeuProAlaLeuAspGluHisArgGluProGlyLeuLeuLeu 397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CTAGAAAATAATAACTGGACTTCCAGCTGTGGATGAACAACCTCAGTTATTGTCTA 420

Qy   398 AspIle-LysGluGluIleGluAspGluGluLysThrIleGluAspTyrIleAspLysLy 417
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GATATTAAAGAAGAAATTGAAGATGAAGAAAAGACAATTGAAGATTATTGATGAAAA 480

Qy   417 sMetAsnAspAlaAspSerThrSerValGlyIleAlaMetTyr-SerValAlaSerGlnCys 436
          ||||||||||||||||||||||||||||||| |||||-|||||||
Db   481 GATGAATGAATGCTTGATTCCACTTCAGTTGAAACTATGTACCTTCGTGTGNTAGTCAAATGT 539

US-10-001-254-16.rng   Page 16

KW   Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW   cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
PS
OS   Homo sapiens.
XX
PN   WO20007301-A2.
XX
DR   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999; 99US-0136526.
PR   10-SEP-1999; 99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PS   Claim 50; Page 303; 799pp; English.
XX
CC   AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC   represent nucleotide sequences encoding human breast, gastric and
CC   prostate cancer associated antigen precursors (CAAP) respectively.
CC   AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC   represent human breast, gastric and prostate CAAP protein sequence
CC   respectively. CAAPs have cytostatic activity and can be used in the
CC   production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC   acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC   condition characterised by expression of an abnormal amount of a protein,
CC   e.g. cancer.
XX
PT   Nucleic acids encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
SQ   Sequence 586 BP; 183 A; 103 C; 120 G; 173 T; 7 other;

US-10-001-254-16 (1-460) x AAF22511 (1-586)

Alignment Scores:
Pred. No.:
Score: 8.07e-81
                          883.00   Length: 586
                                          Matches: 176
Percent Similarity: 97.24%   Conservative: 0
Best Local Similarity: 97.24%   Mismatches: 3
Query Match: 37.01%   Indels: 2
DB: 22   Gaps: 0

Qy   258 LeuCysLeuValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeu 277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   48 CTCTGCTTAGTATATATGTTTACAATGCCTAATGGTTCAATTGCTTAGACAGACTCTTGCTTG 107

Qy   278 AspGlyThrProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsn 297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   108 GATGGTACTCCACCACTTTCTTGGCACATGAGATGCCAAGATTGCTCAGGGTGCCAGCTAAT 167

Qy   298 GlyIleAsnPheLeuHisGluAsnHisHisIleHisArgAspIleTyrSerAlaAsnIle 317
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   168 GGCATCAATTTTCTACATGAAAATCATCATTATTCATAGAGATATTAAAAGTGCAAATATC 227

Qy   318 LeuLeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGlu 337
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   228 TTACTGGATGAAGCTTTTACTGCTAAAATATCTGACTTTGGCCTTGCACGGGCTTCTGAG 287

Qy   338 LysPheAlaGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaPro 357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   288 AAGTTTGCCCAGACAGTCATGACTAGCAGAATTGTGGGAACACAGCTTATATGGCACCA 347

Qy   358 GluAlaLeuArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeu 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 13
AAF22511
ID   AAF22511 standard; cDNA; 586 BP.
XX
AC   AAF22511;
XX
DT   26-MAR-2001 (first entry)
XX
DE   Human breast cancer associated antigen nucleotide sequence SEQ ID NO:90.

```
Db  348 GAAGCTTTCGGTGGAGAAATAACACCCAAATCTGATATTTACACGCTTTGGTGTGGTTTTA 407
Qy  378 LeuGluIleIleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeu 397
Db  408 CTAGAAAATAATAACTGGACTTCCAGCTGTGGATGAACACCGGGAACCTCAGTTATTGCTA 467

Qy  398 AspIle-LysGluGlyGluIleGluAspGluGluLysThrIleGluGluapplysGly 417
Db  468 GATATTAAAAGAAGGAGAAATTGNAGATGAAGAAAGAACAATTGAAGAATTATATTGATT

Qy  417 sMetAsnAspAlaAspSerThrSerValGluAlaMetTyr-SerValAlaSerGlnCys 436
Db  528 GATGAAATGATTGCTGAATTCCACTTCAGTTGAAACTATGTACCTCTTGTTGNTAGTCAATGT 586
```

RESULT 14
AAH09161/c

ID AAH09161 standard; cDNA; 577 BP.
XX
AC AAH09161;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (3'-primer) SEQ ID NO:5996.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3; SEQ ID 5996; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 577 BP; 163 A; 120 C; 105 G; 177 T; 12 other;

Alignment Scores:
Pred. No.:   7.26e-78   Length: 577
Score: 854.00   Matches: 170
Percent Similarity: 93.48%   Conservative: 2
Best Local Similarity: 92.39%   Mismatches: 11
Query Match: 35.79%   Indels: 2
                        22   Gaps: 0
DB:

us-10-001-254-16 (1-460) x AAH09161 (1-577)

```
Qy  219 ValAspIleThrThrGlyLeuGluLeuLysGln-GlnPheAspGlnGluIleLysValMetAl 238
Db  550 GTTGNCANTNNCTACTAAAGACTTGAACCAGCNAGTTTGATCNAGAATN-AAAGTAAATGGC 492

Qy  238 aLysCysGlnHisGluAsnLeuValGluLeuuGlyPheSerSerAspGlyAspAspIle 258
Db  491 AAAGTGTCAACACATGAAAAACTTAGTAGAACTACTTGGTTTCTCAAGTGGATGGAGATGACCTT 432

Qy  258 uCysLeuValTlyValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAs 278
Db  431 CTGCTTAGTATATGTTTACATGCCTAACGGTTCANTGCTAGACAGACACTCTCTTCCTTGGA 372

Qy  278 pGlyThrProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyValAlaAlaAsnGl 298
Db  371 TGGTACTCCACCACTTTCTTGGGCACATGAGATCCAAGATTGCTCAGGGTGCAGCTAATGG 312

Qy  298 YIleAsnPheLeuHisGluAsnHisHisIleHisArgAspHisLysSerAlaAsnIleIle 318
Db  311 CATCAATTTTCTRCATGAAAATCATCATATTCATAGAGATATTAAAAGNGACAAATATCTT 252

Qy  318 uLeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuValAlaArgAlaSerGluLy 338
Db  251 ACTGGATGAAGCTTTTANTGCTAAAATATCTGNCTTTGGCCTTGTCGCACGGGCTTNNGAGA 192

Qy  338 sPheAlaGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGl 358
Db  191 GTTTGCCCAGACAGTCATGACCTAGCAGGAATTGTGGGAACAACAGCTTATATGGCACCAGA 132

Qy  358 uAlaLeuArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLe 378
Db  131 AGCTTTGCGTGGAGAAATAACACCCAAATCTGATATTTACAGCTTTGGTGTGGTTTTACT 72

Qy  378 uGluIleIleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAs 398
Db  71  AGAAAATAATAACTGGACTTCCAGCTGTGGATGAACACCGGGAACCTCAGTTATTGCTAGA 12

Qy  398 pIleLysGlu 401
Db  11  TATTAAAGAA 2
```

RESULT 15
AAV63193

ID AAV63193 standard; cDNA; 483 BP.
XX
AC AAV63193;
XX
DT 25-MAR-2003 (updated)
DT 13-JAN-1999 (first entry)
XX
DE cDNA from clone eh61_1 which encodes a secreted protein.
XX
KW Secreted protein; immune stimulating; suppressing;
KW haematopoiesis regulating activity; tissue growth activity; activin;
KW inhibin activity; chemotactic; chemokinetic activity; haemostatic;
KW thrombolytic activity; anti-inflammatory activity; cadherin;
KW tumour invasion suppressor activity; tumour inhibition activity; de.
XX

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             85..450
FT                   /*tag= a
XX
PN   WO9844113-A1.
XX
PD   08-OCT-1998.
XX
PF   27-MAR-1998;  98WO-US06176.
XX
PR   28-MAR-1997;  97US-0823330.
PR   25-MAR-1998;  98US-0047661.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI   Racie LA, Spaulding V, Treacy M;
XX
DR   WPI; 1998-542703/46.
DR   P-PSDB; AAW80409.
XX
PT   New isolated polynucleotide(s) and secreted proteins - are obtained
PT   from human cDNA libraries prepared from adult testes, foetal brain,
PT   adult brain, adult blood and placenta
XX
PS   Claim 25; Page 82; 124pp; English.
XX
CC   The present sequence encodes a secreted protein. The nucleic acid
CC   sequence is isolated from a human adult blood cDNA library using
CC   probe AAV63204. The polypeptide may have biological activities such
CC   e.g. nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour
CC   inhibition activity or other activities.
CC   (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence 483 BP; 156 A; 84 C; 101 G; 142 T; 0 other;
```

US-10-001-254-16 (1-460) x AAV63193 (1-483)

```
Alignment Scores:
Pred. No.:
Score:                  2.62e-70
Score:                  779.00         Length:        483
Percent Similarity:     98.76%         Matches:       159
Best Local Similarity:  98.76%         Conservative:  0
Query Match:            32.65%         Mismatches:    1
DB:                     19             Indels:        2
                                       Gaps:          0
```

```
Qy    270  LeuLeuAspArgLeuSerCysLeuAspGlyThrProProLeuSerTrpHisMetArgCys  289
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3  TTGCTAGACAGACTTCTTCTTGGATGGTACCCCACCTTTCTTGGACATGAGATGC        62

Qy    290  LysIleAlaGlnGlyAlaAlaGlyIleAsnPheLeuHisGluAsnHisHisIleHis   309
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63  AAGATTGCTCAGGGTGCAGCTAATGGCATTCAATTTCTACATGAAAATCATCATATTCAT  122

Qy    310  ArgAspIleLysSerAlaAlaAsnIleLeuLeuAspGluAlaPheThrAlaLysIleSerAsp  329
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123  AGAGATATTAAAAGTGCAAAATATCTTACTGGATGAAGCTTTTACTGCTAAAATATCTGAC  182

Qy    330  PheGlyLeuAlaArgAlaSerGlyLys-PheAlaGlnThrValMetThrSerArgIleVa  349
                |||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    183  TTTGGCCTTGCACGGGCTTCTGGAAGTTTTGCCCAGACAGTGATGACTAGCAGAGAATTGT  242

Qy    349  1GlyThrThrAlaTyrMetAlaProGluAlaLeuArgGlyGluIleThrProLysSerAs  369
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243  GGGAACAACAGCTTATATGGCACCAGAAGCTTTGCGTGGAGAAATAACACCCAAAATCTGA  302

Qy    369  pIleTyrSerPheGlyValValLeuLeuGluIleIleThrGlyLeuProAlaValAspGl  389
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303  TATTTACAGCTTTGGTGTGGTTTTACTTGAGAAATAATAAACTGGACTTCCAGCTGTGGATGA  362
```

```
Search completed: January 18, 2004, 01:32:15
Job time : 567.702 secs
```

```
Db    303  TATTTACAGCTTTGGTGTGGTTTTACTTGAGAAATAATAAACTGGACTTCCAGCTGTGGATGA  362
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    389  uHisArgGluProGlnLeuLeuLeuAspIleGlyGluAspGluGluLysTh  409
Db    363  ACACCGTGGAACCTCAGTTATTCGTTAGATATTAGAAGAAATTGAAGATGAAGAAAAGAC  422

Qy    409  rIleGluAspTyrIleAspLysLysMetAsnAspAlaAspSerThrSerValGluAlaMe  429
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423  ATT-GAAGATTTATTTGATAAAAAGATGAATGATGCTGATTCCACTTCAGTTGAAGCTAT  481

Qy    429  t  429
Db    482  G  482
```

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 01:04:18 ; Search time 7784.21 Seconds
(without alignments)
2417.512 Million cell updates/sec

Title: US-10-001-254-16
Perfect score: 2386
Sequence: 1 MNKPITPSTYVRCLNVGLIR...........KNKRPDIKKVQQLLQEMTAS 460

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152423_19714/app_query.fasta_1.1109
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10001254 @CGN 1 1 5671 @runat_16012004_152423_19714 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -FGAPOP=6

Database :　GenEmbl:*

```
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_vr:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2386	100.0	1383	6	AX431306	AX431306 Sequence
	2	2386	100.0	1629	9	BC013316	BC013316 Homo sapi
	3	2386	100.0	2820	6	AK000528	AK000528 Homo sapi
	4	2382	99.8	1383	6	AX196260	AX196260 Sequence
	5	2382	99.8	1383	9	AF445802	AF445802 Homo sapi
	6	2362	99.0	2817	9	AX431318	AX431318 Sequence
	7	2362	99.0	2817	9	AF155118	AF155118 Homo sapi
	8	2017	84.5	1542	10	AX196262	AX196262 Sequence
	9	2017	84.5	1542	10	AF445803	AF445803 Mus muscu
	10	2016	84.5	2431	6	BC051676	BC051676 Mus muscu
c	11	1787.5	74.9	2213	9	BD155790	BD155790 Primer fo
	12	1787.5	74.9	2213	9	AK027301	AK027301 Homo sapi
	13	1169	49.0	833	9	AR223870	AR223870 Sequence
	14	981	41.1	1719	5	BC045381	BC045381 Danio rer
	15	883	37.0	539	6	AX053431	AX053431 Sequence
	16	883	37.0	586	6	AX053324	AX053324 Sequence
c	17	854	35.8	577	6	BD151153	BD151153 Primer fo
	18	691	29.0	501	6	AX321132	AX321132 Sequence
	19	666	27.9	2069	3	AK116443	AK116443 Ciona int
	20	590	24.7	33033	9	AY186092	AY186092 Homo sapi
c	21	590	24.7	118572	2	AC093012	AC093012 Homo sapi
c	22	590	24.7	165868	2	AC021719	AC021719 Homo sapi
c	23	590	24.7	309582	2	AC025567	AC025567 Homo sapi
	24	574	24.1	1287	8	AX653723	AX653723 Homo sapi
c	25	528.5	22.2	66237	8	AB016892	AB016892 Arabidops
	26	525.5	22.0	2643	8	BT000634	BT000634 Arabidops
	27	525.5	22.0	2821	8	AY091785	AY091785 Arabidops
	28	517	21.7	294	6	AX431296	AX431296 Sequence
	29	515.5	21.6	1930	8	AY056294	AY056294 Arabidops
	30	515.5	21.6	3590	6	AR270530	AR270530 Sequence
	31	515.5	21.6	3590	6	AX330494	AX330494 Sequence
	32	515.5	21.6	3590	6	AX409575	AX409575 Sequence
	33	515.5	21.6	3590	9	I59730	I59730 Sequence 1
c	34	515.5	21.6	244696	2	L76191	L76191 Homo sapien
c	35	512.5	21.5	252400	2	AC123081	AC123081 Rattus no
	36	512.5	21.5	2361	6	AC094859	AC094859 Rattus no
	37	508.5	21.3	1899	6	AX506978	AX506978 Sequence
	38	506	21.2	1930	6	AX506104	AX506104 Sequence
	39	506	21.2	2585	8	AY150495	AY150495 Arabidops
	40	506	21.2	2815	8	AY224496	AY224496 Arabidops
	41	506	21.2	1281	8	AX412672	AX412672 Oryza sat
	42	505.5	21.2	1281	8	AY125559	AY125559 Sequence
	43	505.5	21.2	1539	8	D88207	D88207 Arabidopsis
	44	505.5	21.2	1588	8	AF428432	AF428432 Arabidops
	45	505	21.2	3051	8	BT002512	BT002512 Arabidops

RESULT 1

ALIGNMENTS

```
LOCUS       AX431306              1383 bp    DNA     linear   PAT 28-JUN-2002
DEFINITION  Sequence 15 from Patent WO0240680.
ACCESSION   AX431306
VERSION     AX431306.1  GI:21656175
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
            Roth,W. and Stenner-Liewen,F.
  TITLE     Novel death domain proteins
  JOURNAL   Patent: WO 0240680-A 15 23-MAY-2002;
            BURNHAM INST (US)
FEATURES             Location/Qualifiers
     source          1..1383
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             1..1383
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD37279.1"
                     /db_xref="GI:21656176"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDPIDPQEGWKKLAVAIKKPSG
                     DDRNQFHIRRFEALLQTGKSPTSELLFDWGTVNCTVGDLVDLLIQNEFPAPALLLP
                     DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMPPVQNLEQSVMPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTEELKQQFDQEIKVNAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRISC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHIIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTPAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLLDIKEEIEDEEKTIEDYIDKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                     CQLLQEMTAS"
BASE COUNT      463 a    243 c    283 g    394 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                      1
Score:                  2386.00
Percent Similarity:    100.00%
Percent Similarity:    100.00%
Best Local Similarity: 100.00%
Query Match:                 6
DB:                          6
```

```
US-10-001-254-16 (1-460) x AX431306 (1-1383)
```

```
Qy          1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg   20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAAATGTTGGACTAATTAGG   60

Qy         21 LysLeuSerAspProIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys   40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA  120

Qy         41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu   60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGATTTGAAGCATTACTT  180

Qy         61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr   80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCACACAAATTGCACA  240

Qy         81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPheAlaProAlaSerLeuLeuLeu  100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGGTGATCTTGTTGATCTTTTGATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTG  300

Qy        101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTCCCAGATGCTGTTCCAAAACTGCTAATACACTACCTTCTAAAGAAGCTATAACAGTT  360

Qy        121 GlnGlnLysGlnMetProPheCysAspLysAspArgThrLeuMetProProValGlnAsn  140
```

```
Db        361 CAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGACATTGATGACACCTTGATGCCACCTGTTCAGAAT  420

Qy        141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal  160
Db        421 CTTGAACAAAGCTATATGCCACCTGACTCCTCAAGTCCAGAAAAATAAAAGTTTAGAAGT  480

Qy        161 SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe  180
Db        481 AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGAAGAATGTCACAAATAACTTT  540

Qy        181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal  200
Db        541 GATGACGACCCCATTTCTGTTGGTGGTAATAAAATGGGAGAGGGAGGATTTGGAGTTGTA  600

Qy        201 TyrLysGlyTyrValAsnAsnThrThrValAlaValLysLysLeuAlaAlaMetValAsp  220
Db        601 TATAAAGGCTACGTAAATAACACAACTGTGGCAGTGAAGAAGCTTGCAGCCAATGGTGAC  660

Qy        221 IleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysValAsnAlaLysCys  240
Db        661 ATTACTACTGAAGAACTGAAACAGCAGTTTGATCAAGAAATAAAGGTCAACGTGAAGT  720

Qy        241 GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu  260
Db        721 CAACATGAAAAACTTAGTAGAACTACTTGGTTTCTCAAGTGATGGAGATGACCTCTGCTTA  780

Qy        261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgIleSerCysLeuAspGlyThr  280
Db        781 GTATATGTTTACATGCCTAATGGTTCATTGCTAGACAGAATCTCTTGCCTTGGATGGTACT  840

Qy        281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn  300
Db        841 CCACCACTTCTTGGCACATGAGATGCCAAGATTGCTCAGGGTGCAGCTAATGGCATCAAT  900

Qy        301 PheLeuHisGluAsnHisIleIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp  320
Db        901 TTTCTACATGAAAATCATATTATTCATAGAGATATTAAAAGTGCAAATATCTTACTTGAT  960

Qy        321 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla  340
Db        961 GAAGCTTTTACTGCTAAAATATCTGACTTTGGCCTTGCAAGGGCTTCTGAGAAGTTTGCC 1020

Qy        341 GlnThrValMetThrSerArgIleValGlyThrProAlaTyrMetAlaProGluAlaLeu  360
Db       1021 CAGACAGTTCATGACTAGCAGAATTGTGGGAACAGCCTATATGGCACCAGAAGCTTTG 1080

Qy        361 ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle  380
Db       1081 CGTGGGAGAAATAACACCCAAATCTGATATTTACAGCTTTGGTGTGGTTTTACTAGAAATA 1140

Qy        381 IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAspIleLys  400
Db       1141 ATAACTGGACTTCCAGCTGTTGATGAACACCGTGAACCTCAGTTATTGCTAGATATTAAA 1200

Qy        401 GluGluIleGluAspGluGluLysThrIleGluAspTyrIleAspLysMetAsnAsp  420
Db       1201 GAAGAAATTGAAGATGAAGAAAAGACAATTGAAGATTATATTGATAAGATGAATGAT 1260

Qy        421 AlaAspSerThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys  440
Db       1261 GCTGATTCCACTTCAGTTGAAGCTATGTACTCTGTTGCTAGTCAGTGCCTGCATGAAAAG 1320

Qy        441 LysAsnLysArgProAspIleLysLysValGlnLeuLeuGlnGluMetThrAlaSer  460
Db       1321 AAAAATAAGAGACCAGACATTAAGAAGGTTCAACAGCTGCTCCAAGAGATGACAGCTTCT 1380
```

```
RESULT 2
BC013316
LOCUS       BC013316              1629 bp    mRNA    linear   PRI 04-SEP-2001
DEFINITION  Homo sapiens, clone MGC:13330 IMAGE:4287014, mRNA, complete cds.
ACCESSION   BC013316
VERSION     BC013316.1  GI:15426431
```

```
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1629)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 19 Row: n Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7705840.
FEATURES             Location/Qualifiers
     source          1..1629
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:13330 IMAGE:4287014"
                     /tissue_type="Brain, primitive neuroectodermal"
                     /clone_lib="NIH_MGC_56"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             71..1453
                     /codon_start=1
                     /product="Unknown (protein for MGC:13330)"
                     /protein_id="AAH13316.1"
                     /db_xref="GI:15426432"
                     /translation="MNKPIPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRVNQFHIRRFEALLQTGKSPSTSELLFDWGTTNCTVGDLVDLLIQNEFPAPASLLP
                     DAVRTANYLPSKEAITVQQKOMPFCDKDRTLMTPVVQNLEQSYMPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTIVAVKKLAAM
                     VDITTEELKQQFDQEIKVWAKCQHENLVELLGFSSGDDLCLVVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     QOLLDIEIEDEEKTIEDYIDKKWNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                     QOLLQEMIAS"
BASE COUNT        542 a        291 c        328 g        468 t
ORIGIN
```

```
Qy      1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAACAAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 60

Qy     21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGCTCTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 120

Qy     41 LysProSerGlyAspAspArgValAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAACCATCTGGTGATGATAGAGTCAACCAGTTTCACATAAGGAGATTTGAAGCATTACTT 180

Qy     61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGACCACAAATTGCACA 240

Qy     81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 GTTGGTGATCTTGTGGATCTTTTGATCCAAAATGAATTTTTTGCTCCTGCAAGTCTTTTG 310

Qy    101 ProAspAlaValArgThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 CCAGATGCTGTTCGCACAGCCAACACTCTACCCTCAAAAGAAGCAATAACAGTT 370

Qy    121 GlnGlnLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 CAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGGACATTGATGACACCTGTGCAGAAT 430

Qy    141 LeuGluGlnSerTyrMetProAspSerSerSerProGluAsnLysSerLeuGluVal 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    431 CTTGAACAAAGCTATATGCCAGACTCCTCAAGTCCAGAAAATAAAAGTTTAGAAGTT 490

Qy    161 SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    491 AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGAAAAATGTCACAAATAACTTT 550

Qy    181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551 GATGAACGACCAATTTCTGTTGGTGGTAATAAAATGGGAGAAGGTGGATTTGGAGTTGTA 610

Qy    201 TyrLysGlyTyrValAsnAsnThrThrValAlaValLysLysLeuAlaAlaMetValAsp 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 TATAAAGGCTACGTAAATAACACAACTGTGGCTGTGAAAAAGCTTGCAGCAATGGTTGAC 670

Qy    221 IleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysValTrpAlaLysCys 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 ATTACTACTGAGGAACTAAAACAGCAGTTTGATCAAGAAATAAAAGTGGCATGGAAGTGT 730

Qy    241 GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAlaGlyAspAspLeuCysLeu 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 CAACATGAAAATTTAGTAGAACTACTTGGTTTTTCTAGTGCTGGTGATGACCTCTGCTTA 790

Qy    261 ValValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 GTTGTATATATGCCTAATGGTTCACTGCTTGATCGACTTTCATGTCTTGGATGGTACT 850

Qy    281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 CCACCACTTCTTGGCACATGAGATGCGTTGCAAGATTGCTCAGGGTGCAGCTAATGGCAT 910

Qy    301 PheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 TTTCTACATGAAAATCATCATATTCATAGAGATATTAAAAGTGCAAATATCTTACTGGAT 970

Qy    321 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 GAAGCTTTTACTGCTAAAATATCTGACTTTGGCCTTGCCAGAGCTTCTGAGAAGTTTGCC 1030

Qy    341 GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 CAGACAGTTATGACTAGCAGAATTGTGGGAACAACAGCTTATATGGCACCAGAAGCTTTG 1090

Qy    361 ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 CGTGGAGAAATAACACCCAAAAGTGATATTTACAGCTTTGGGGTTGTTTTACTAGAAATA 1150

Qy    381 IleThrGlyLeuProAlaValAspGluHisArgGluGlnLeuLeuAspIleLys 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 ATAACTGGACTTCCAGCTGTGGATGAACACCGTGAACCTCAGTTATTGGATATTAAA 1210

Qy    401 AsnHisProSerThrPheAsnThrTyr... 
Db   1211 ATAACTGGACTTCCAGCTGTGGATGAACACCGTGAACCTCAGTTATTGGATATTAAA 1270
```

```
Qy    401  GluGluIleGluAspGluGluLysThrIleGluValAspTyrIleAspGlyProAspTyrIleAspGly  420
Db   1271  GAAGAAATTGAAGATGAAGAAAAGACAATTGAAGTAGATTATATTGATGGAAAGATGAATGAT  1330

Qy    421  AlaAspSerThrSerValGluValAlaMetTyrSerValAlaSerGlnCysGluHisGlyLys  440
Db   1331  GCTGATTCCACTTCAGTTGAAGCTATGTACTCTGTTGCTAGTCAGTGTGAACATGGAAAG  1390

Qy    441  LysAsnLysArgProAspIleLysLysValIleGlnIleLeuLeuGlnGluMetThrAlaSer  460
Db   1391  AAAAATAAGAGACCAGACATTAAGAAGGTTCAACAGCTGCTGCAAGAGATGACAGCTTCT  1450
```

RESULT 3

LOCUS   AK000528

DEFINITION   Homo sapiens cDNA FLJ20521 fis, clone KAT10395.

ACCESSION   AK000528

VERSION   AK000528.1   GI:7020683

KEYWORDS   oligo capping; fis (full insert sequence).

SOURCE   Homo sapiens (human)

ORGANISM   Homo sapiens
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (sites)

AUTHORS   Watanabe,S., Kunagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
     Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Isogai,T. and Sugano,S.
     Nakamura,Y.

TITLE   NEDO human cDNA sequencing project

JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 2820)

AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
     Shibahara,T., Tanaka,T. and Nakamura,Y.

TITLE   Direct Submission

JOURNAL   Submitted (15-FEB-2000) Sunio Sugano, Institute of Medical Science,
     University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1,
     Minato-ku Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp,
     Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT   NEDO human cDNA sequencing project supported by Ministry of
     International Trade and Industry of Japan; cDNA full insert
     sequencing: Research Association for Biotechnology; cDNA library
     construction, 5'- & 3'-end one pass sequencing: Department of
     Virology and Human Genome Center, Institute of Medical Science,
     University of Tokyo (partly supported by Science and Technology
     Agency).

FEATURES   Location/Qualifiers

source   1..2820
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="KAT10395"
     /cell_line="KATO III"
     /cell_type="signet-ring cell carcinoma"
     /clone_lib="KAT"
     /note="cloning vector pME18SFL3"

CDS   30..1412
     /note="unnamed protein product"
     /codon_start=1
     /protein_id="BAA91232.1"
     /db_xref="GI:7020684"
     /translation="MNKPITPSTYVRCLNVGLIRKLSDFIPDQEGWKKLAVAIKKPSG
     DDRYNQFHIRFERALLQTGKSPTSELILFDWDTNTCTVCDLVDLLIQNEFFAPASLILP
     DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSYMPDSSSPENKSLEV
     SDTRFHSFSFYELKNVTNNFDERPISVGGGNKMGEGGFGVVYKGVVNNTTVAVKKLAAM
     VDITTEELKQQFDQEIKVMAKCQHENLVELIGFSSDGDDLCLVYVPMNGSLLDRLSC
     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDISANILLDEAFTAKISDFGIAR
     ASEKFAQVTWMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVLLEIITGLPAVDEHRE
     PQLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMTSVASQCLHEKNKRPDIKKV
     QLLLQEMTAS"

BASE COUNT   940 a   534 c   579 g   767 t

ORIGIN
```

(handwritten annotations: "DNA encoding Seq ID #16", "2820 bp", "PRT 22-FEB-2000", AK000528, Watanabe)

```
LOCUS       AK000528      2820 bp    mRNA    linear
DEFINITION  Homo sapiens cDNA FLJ20521 fis, clone KAT10395.
```

Alignment Scores:

Pred. No.:   1.26e-197   Length: 2820

US-10-001-254-16 (1-460) x AK000528 (1-2820)

Score: 2386.00   Matches: 460

Percent Similarity: 100.00%   Conservative: 0

Best Local Similarity: 100.00%   Mismatches: 0

Query Match: 9   Indels: 0

DB: Gaps: 0

```
Qy     1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeu  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    30  ATGAACAAACCCATAACAACCATCAACATATGTGCGCTGCCTCAATGTGGGACTT  89

Qy    21  LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLysLys  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    90  AAGCTGTCAGATTTTATTGATCCTCAAGAAGGGATGGAAGAAGTTAGCTGTAATAAAA  149

Qy    41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgPheGluArgAlaLeuLeu  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   150  AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGATTTGAACGATACTT  209

Qy    61  GlnThrGlyLysSerProThrSerGluLeuIleLeuPheAspTrpGlyThrThrAsnCysThr  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210  CAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCACCACAAATTGCACA  269

Qy    81  ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePhePheAlaProAlaSerLeuLeu  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   270  GTTGGTGATCTTGTGGATCTTTTGATCCAAAATGAATTTTTTGCTCCCGCAAGTCTTTTG  329

Qy   101  LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   330  CTCCCAGATGCTGTTCCCAAAACTGCTAATACACTACCTTCTAAAGAAGCTATAACAGTT  389

Qy   121  GlnGlnLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   390  CAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGAACATTGATGACACCTGTGCAGAAT  449

Qy   141  LeuGluGlnSerTyrMetProAspSerSerSerProGluAsnLysSerLeuGluVal  160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   450  CTTGAACAGAAGCTATATGCCAGACTCTCCAGTCCAGAAAATAAAAGTTTAGAAGTT  509

Qy   161  SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   510  AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGAAGAATGTCACAAATAACTTT  569

Qy   181  AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal  200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   570  GATGAACGACCCCATTTCTGTTGGTGGTAATAAAATGGGAGAGGGAGGATTTGGAGTTGTA  629

Qy   201  TyrLysGlyValValAsnAsnThrThrValAlaValLysLysLeuAlaAlaMetValAsp  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   630  TATAAAGGCGTAGTGAATAACACTACAGTTGCTGTAAAGAAGCTTGCAGCAATGGTTGAC  689

Qy   221  IleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysValMetAlaLysCys  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   690  ATTACTACTGAAGAACTCAAACAGCAGTTTGATCAAGAAATAAAAGTGATGGCAAAGTGT  749

Qy   241  GlnHisGluAsnLeuValGluLeuIleGlyPheSerSerAspGlyAspAspLeuCysLeu  260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   750  CAACACGAAAACTTAGTAGAACTCATTGGTTTCTCAAGTGATGGGGATGACCTCTGCTTA  809

Qy   261  ValTyrValProMetAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr  280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   810  GTATATGTTTACATGCCTAATGGTTCATTGCTAGACCGACTCTCTTGCCTGGATGGTACT  869

Qy   281  ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   870  CCACCACTTCTGGGCACATGAGATGCCAAGATTGTCTCAGCCTGCACATGGCATCAAT  929

Qy   301  PheLeuHisGluAsnHisHisIleHisArgAspIleSerAlaAsnIleLeuLeuAsp  320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   930  TTTCTACATGAAAATCATCATATTCATAGAGATATATCAGCAAATATTCTTACTCGAT  989

Qy   321  GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla  340
```

```
BASE COUNT      463 a     244 c     283 g     393 t
ORIGIN

Alignment Scores:
Pred. No.:              1.11e-197        Length:          1383
Score:                  2382.00          Matches:         459
Percent Similarity:     99.78%           Conservative:    0
Best Local Similarity:  99.78%           Mismatches:      1
Query Match:            99.83%           Indels:          0
```

```
RESULT 4
LOCUS         AX196260              1383 bp      DNA        linear     PAT  28-AUG-2001
DEFINITION    Sequence 2 from Patent WO0151641.
ACCESSION     AX196260
VERSION       AX196260.1   GI:15386462
KEYWORDS      .
SOURCE        Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1
  AUTHORS     Wesche,H. and Li,S.
  TITLE       Irak-4: compositions and methods of use
  JOURNAL     Patent: WO 0151641-A 2 19-JUL-2001;
              Tularik Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1383
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="human IL-1 receptor-associated kinase 4 (IRAK-4)
                     cDNA."
     CDS             1..1383
                     /note="human IRAK-4"
                     /codon_start=1
                     /protein_id="CAC60090.1"
                     /db_xref="GI:15386463"
                     /translation="MNKPIFPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRFEALLQTGKSPTSELLFDWGTTNCTAGDLVDLLIQNEFFAPASLLP
                     DAVPKTANTLPSKEAITVQOKOMPFCDKDRTLMTPVQNLEQSYMPDSSSPENKSLEV
                     SDTRHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTEELKQQFDQEIKVMAKCQHENIVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     POLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNRPDIKKV
                     QQLLQEMTAS"
```

```
DB:             US-10-001-254-16 (1-460)   x AX196260 (1-1383)
                                    6             Gaps:        0

Qy      1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  ATGAACAAACCCATAACACCATCAACACATTATGTGCGGTGCCTCAATGTTGGACTAATTAGG  60
```

... (sequence alignment continues in paired rows of Query (Qy) and Database (Db) nucleotide/protein sequences through positions to 1383 / 460)

```
RESULT 5
LOCUS       AF445802    1383 bp    mRNA    linear    PRI 20-APR-2002
DEFINITION  Homo sapiens interleukin-1 receptor associated kinase 4 (IRAK4)
            mRNA, complete cds.
ACCESSION   AF445802
VERSION     AF445802.1  GI:20219009
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1383)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE     IRAK-4: a novel member of the IRAK family with the properties of an
            IRAK-kinase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (8), 5567-5572 (2002)
  MEDLINE   21957277
   PUBMED   11960013
REFERENCE   2  (bases 1 to 1383)
  AUTHORS   Suzuki,N., Suzuki,S., Duncan,G.S., Millar,D.G., Wada,T.,
            Mirtsos,C., Takada,H., Wakeham,A., Itie,A., Li,S., Penninger,J.M.,
            Wesche,H., Ohashi,P.S., Mak,T.W. and Yeh,W.C.
  TITLE     Severe impairment of interleukin-1 and Toll-like receptor
            signaling in mice lacking IRAK-4
  JOURNAL   Nature 416 (6882), 750-756 (2002)
  MEDLINE   21959395
   PUBMED   11923871
REFERENCE   3  (bases 1 to 1383)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) Biology I, Tularik Inc., 2 Corporate Drive,
            South San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1..1383
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..1383
                     /gene="IRAK4"
     CDS             1..1383
                     /gene="IRAK4"
                     /note="protein kinase"
                     /codon_start=1
                     /product="interleukin-1 receptor associated kinase 4"
                     /protein_id="AAM15772.1"
                     /db_xref="GI:20219010"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQIGKSPTSELLFDWGTTNCTAGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSYMPPDSSSPENNSLEV"
```

```
BASE COUNT      463 a    244 c    283 g    393 t
ORIGIN

                                                    SDTRFHSFSFYELKNVTNNPDERPISVGGNKMGEGGFGVVVKGYVNNTTVAVKKLAAM
                                                    VDITTEELKQQFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVMPNGSLLDRLSC
                                                    LDGTPPLSWHMRCKIIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                                                    ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVLLLEIITGLPAVDEHRE
                                                    PQLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIIKV
                                                    QQLLQEMTAS"
```

```
Alignment Scores:
Pred. No.:
Score:                    1.11e-197       Length:         1383
Percent Similarity:       2382.00         Matches:        459
Best Local Similarity:    99.78%          Conservative:   0
Query Match:              99.78%          Mismatches:     1
DB:                       99.83%          Indels:         0
                          9               Gaps:           0
```

```
US-10-001-254-16 (1-460) x AF445802 (1-1383)

Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGly 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGAACAAACCCATAACAACCATCAACATATGTCGGTGCCTCAATGTGGGT 60

Qy    21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLysLys 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 120

Qy    41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGAATTGAAGCATTACTT 180

Qy    61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCACCACAAATTGCACA 240

Qy    81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeu 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCTGGTGATCTTGTTGATCTTTTGATCCAAAATGAATTTTTTGCTCCTGCAAGTCTTTTG 300

Qy   101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTCCCAGATGCTGTTCCCAAAACTGCTAATACACTACCTTCTAAAGAAGCTATAACAGTT 360

Qy   121 GlnGlnLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGAACATTGATGACACCTGTGCAGAAT 420

Qy   141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnAsnSerLeuGluVal 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CTTGAACAAAGCTATATGCCACCTGACTCCTCAAGTCCAGAAAATAAAAGTTTAGAAGTT 480

Qy   161 SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGAAGAATGTCACAAATAACTTT 540

Qy   181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GATGAACGACCCATTTCTGTTGGTGGTAATAAAATGGGAGAGGGAGGATTTGGAGTGGTA 600

Qy   201 TyrLysGlyTyrValAsnAsnThrThrValAlaValLysLysLeuAlaAlaMetValAsp 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TATAAAGGCTACGTAAATAACACAACTGTGGCAGTGAAGAAGCTTGCAGCAATGGTTGAC 660

Qy   221 IleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysValMetAlaLysCys 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATTACTACTGAAGAACTGAAACAGCAGTTTGATCAAGAAATAAAAGTAATGGCAAAGTGT 720

Qy   241 GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu 260
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 CAAACATGAAAACTTAGTAGAACTACTTGGTTTCTCAAGTGATGGAGATGACCTCTGCTTA 780

Qy   261 ValTyrValMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr 280
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  1021 CAGACAGTCATCAACTACCAGAATTGTGGGACAACAGCTTATATGGCACCAGAAGCTTTG 1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle 380

Qy  1081 CGTTGGAGAAAATAACACCCAAATCTGATPATTTTACAGCTTTGGTGTGGTTTTTACTAGG 1140
         |||||||||||||||||||||||||||||||
Db   381 IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAspIleLys 400

Qy  1141 ATAACTGGACTTCCAGCTGTGGATGAACATAGGGAACCTCAGCTTTTGTTGGATATAAAA 1200
         ||||||||||||||||||||||||||||||||||
Db   401 GluGluIleGluAspGluGluLysThrIleGluAspTyrIleAspLysLysMetAsnAsp 420

Qy  1201 GAAGAAATTGAAGATGAAGAAAAAGACAAATTGAAGATTATATGATAAAGAAGATGAATGAT 1260
         ||||||||||||||||||||||||||||||||||||||
Db   421 AlaAspSerThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys 440

Qy  1261 GCTGATTCACTTCAGTTGAAGCTATGTACCTCGTTGCTAGTCAAGTGCTTCCATGAAAAG 1320
         ||||||||||||||||||||||||||||||||||||||
Db   441 LysAsnLysArgProAspIleIleLysValGlnGlnLeuLeuGlnGluMetThrAlaSer 460

Qy  1321 AAAAAATAAGAGACCAGAACATTAAGAAGGTTCAACAGGTGCTGCAAGAGATGACAGCTTCT 1380
         ||||||||||||||||||||||||||||
```

RESULT 6

AX431318

```
LOCUS       AX431318      2817 bp    DNA      linear   PAT 28-JUN-2002
DEFINITION  Sequence 27 from Patent WO0240680.
ACCESSION   AX431318
VERSION     AX431318.1  GI:21656187
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
            Roth,W. and Stenner-Liewen,F.
  TITLE     Novel death domain proteins
  JOURNAL   Patent: WO 0240680-A 27 23-MAY-2002;
            BURNHAM INST (US)
FEATURES             Location/Qualifiers
     source          1..2817
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             50..1432
                     /codon_start=1
                     /protein_id="CAD37285.1"
                     /db_xref="GI:21656188"
                     /note="unnamed protein product"
                     /translation="MNKPIPSTYVRCLNVGLIRKLSDFIDPQBGWKKLAVAIKPSG
                     DDRVNQFHIRRFEALLQTGKSPTSELLFDWGTLVDILLQNEFPAPASLLLP
                     DAVPKTANTLPSKEAITVQQKQMPFCDKDTLMTPVQNLEQSYMPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNVMGEGGFGVVYKGYVNNTTVAVKLLAAM
                     VDITEELKQOFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     POLLLDIKEIEDEKTIEDYIDKKWNDADSTSVEAMYSGASQCRHEKKNKSPDIKKV"
BASE COUNT   912 a   547 c   586 g   772 t
ORIGIN
```

US-10-001-254-16 (1-460) x AX431318 (1-2817)

```
Alignment Scores:
Pred. No.:
Score:                   1.54e-195     Length:        2817
Percent Similarity:      99.13%        Matches:       456
Best Local Similarity:   99.13%        Conservative:  0
Query Match:             98.99%        Mismatches:    4
DB:                                    Indels:        0
                                       Gaps:          0
```

```
Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
Db    1 ATGAACAAACCCATAACACCATCAACATATGTGCGGTGCCTCAATGTTGGACTAATTAGG 50

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys 40
Db   50 AAGCTGTCAGATTTTATTGATCCTCAAGAGGGATGGAAGAAGTTAGCTGTAGCTATTAAA 109

Qy   41 LysProSerGlyAspAspArgValAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
Db  110 AAACCATCTGGTGATGATAGAGTCAATCAGTTTCACATAAGGAGGTTTGAAGCATTACTT 169

Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrLeuAsnCysThr 80
Db  170 CAAACTGGAAAAAGTCCCACTTCAGAATTACTGTTTGACTGGGGAACAAATTGCACA 229

Qy   81 ValGlyAspLeuValAspIleLeuIleGlnAsnGluPheProAlaSerLeuLeu 100
Db  230 GTTGGTGATCTTGTGGACATCTTTTTGATCCAAAATGAATTTTTGCCTCCAGTCTTTTG 289

Qy  101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120
Db  290 CTCCCAGATGCTGTTCCAAAACTGCTAATACTACCTTCTAAAGAAGCTATAACAGTT 349

Qy  121 GlnGlnLysGlnMetProPheCysAspLysAspThrLeuMetThrProValGlnAsn 140
Db  350 CAGCAAAAACAGATGCCTTTCTGTGACAAAGACACATTGATGACACCTGTGCAGAAT 409

Qy  141 LeuGluGlnSerTyrMetProAspSerSerSerProGluAsnLysSerLeuGluVal 160
Db  410 CTTGAACAAAGCTATATGCCACTGACTCCTCAAGTTCCAGAAAATAAAGTTTAGACTT 469

Qy  161 SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe 180
Db  470 TATAAAAGGCTACCTAAATAACACAACTGTGGACAGTGAAGAAGTTCAGCAAGGTTCAC 529

Qy  181 AspGluArgProIleSerValGlyGlyAsnValMetGlyGluGlyGlyPheGlyValVal 200
Db  530 AGTGATACACGTTTTCACAGTTTTCATTTTGTAGGAATGTCCAGAATAACTT 589

Qy  201 TyrLysGlyTyrValAsnAsnThrThrValAlaValLysLeuLeuAlaAlaMetValAsp 220
Db  590 GATGAACGGACCCATTTCTGTTGGTGGTAATAAATGGGAGGGAGGAGAATTGGAGTTGTA 649

Qy  221 IleThrGluGluLeuLysGlnPheSerSerAspGlyAspAspLeuCysLeu 240
Db  650 ATTACTACTGAAGAACTGAAACAGCAGTTTGATCAAGAAATAAAAGTAATGGCAAAGTGT 709

Qy  241 GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu 260
Db  710 CAACATGAAAACTTAGTAGAACACTTAGGTTTTTCAAGTGATGGAGATGATCTCTGCTTA 769

Qy  261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr 280
Db  770 GTATATGTTTACATGCCTAATGGTTCATTGCTAGACAGACTCTCTTGCTTGGATGGTACT 829

Qy  281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn 300
Db  830 CCACCACTTTCTTGGCACATGAGATGCAAGATTGCTCAAGGTGCCAATGGCATCAAT 889
```

RESULT 7

AF155118

LOCUS   AF155118   2817 bp   mRNA   linear   PRI 05-JAN-2000
DEFINITION   Homo sapiens putative protein kinase NY-REN-64 antigen mRNA,
             complete cds.
ACCESSION   AF155118
VERSION   AF155118.1 GI:5360130
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2817)
  AUTHORS   Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
             Jongeneel,V., Gure,A.O., Jager,D., Jager,E., Knuth,A., Chen,Y.-T.
             and Old,L.J.
  TITLE   Antigens recognized by autologous antibody in patients with
             renal-cell carcinoma
  JOURNAL   Int. J. Cancer 83 (4), 456-464 (1999)
   MEDLINE   99438124
   PUBMED   10508479
REFERENCE   2 (bases 1 to 2817)
  AUTHORS   Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
             Jongeneel,V., Gure,A.O., Jager,D., Jager,E., Knuth,A., Chen,Y.-T.
             and Old,L.J.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Ludwig Institute, Sloan-Kettering
             Institute, 1275 York Ave, New York, NY 10021, USA
FEATURES   Location/Qualifiers
     source   1..2817
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="1973/10-4"
                     /cell_type="renal cell carcinoma"
     CDS   50..1432
                     /codon_start=1
                     /product="putative protein kinase NY-REN-64 antigen"
                     /protein_id="AAD42884.1"
                     /db_xref="GI:5360131"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDPIDPQEGWKKLAVAIKKPSG
                     DDRNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVGQKMPFCDKDRTLMTPVQNLEQSYMPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNFDERPISVGGNKMGEGGFGVVYKGVVNNTVAVKKLAAM
                     VDITTEELKQQFDQEIKVWAKCQHENLVELLGFSSDGDDLCLVVYMPNGSLLDRLSC
                     LDGTPPLSWHMRKKIAQGAANGINFLHNHHIRRDIKSANILLDEAFTAKISDFGLAR
                     ASERFAQTVMTSRIVGTYAYMAPEALRGEITTKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLDKIEEIEDEKTIEDYIDKKMNDADSTSVEAMYSGASQCRHEKNKSPDIKKV
                     HQLLQEMTAS"

BASE COUNT   912 a   547 c   586 g   772 t
ORIGIN

US-10-001-254-16 (1-460) x AF155118 (1-2817)

Alignment Scores:
Pred. No.:
Score: 1.54e-195   Length: 2817
Score: 2362.00   Matches: 456
Percent Similarity: 99.13%   Conservative: 0
Best Local Similarity: 99.13%   Mismatches: 4
Query Match: 98.99%   Indels: 0
DB: 9   Gaps: 0

Qy   1   MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg   20
Db   1   ATGAACAAACCCATAACACCATCAACATACGTGCGCTGCCTTCAAATGTTGGACTAATTAGG   109

Qy   21   LysLeuSerAspProIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110   AAGCTGTCAGATTTTATTGATCCTCAAGAGGGATGGAAGAAGTTAGCTGTAGCTATTAAA   169

Qy   41   LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 170   AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGAATTTGAAGCATTACTT   229

Qy   61   GlnThrGlyLysSerProThrSerGluLeuLeuLeuPheAspTrpGlyThrThrAsnCysThr   80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 230   CAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTGTTTGACTGGGGAACAACATTGCACA   289

Qy   81   ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeu   100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 290   GTTGGTGATCTTGTGGATCTTTTGATCCAAAATGAATTTTTTGCTCCTGCAAGTCTTTTG   349

Qy 101   LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal   120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 350   CTCCCAGATGCTGTTCCCAAAACTGCTAATACACTGCCTTCTAAAGAAGCTATAACAGTT   409

Qy 121   GlyGlnLysMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn   140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 410   GGGCAAAAAATGCCTTTCTGTGACAAAGACAGGACAATTGATGACACCTGTGCAGAAT   469

Qy 141   LeuGluGlnSerTyrMetProAspSerSerSerProGluAsnLysSerLeuGluVal   160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 470   CTTGAACAAAGCTATATGCCAGACTCCTCAAGTCCAGAAAATAAAAGTTTAGAAGTT   529

Qy 161   SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe   180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 530   AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGAAGAATGTTACAAATAACTTT   589

Qy 181   AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal   200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 590   GATGAACGACCAATTTCTGTTGGTGGTAATAAAATGGGAGAGGGAGGATTTGGAGTTGTA   649

Qy 201   TyrLysGlyValValAsnAsnThrValAlaValLysLysLeuAlaAlaMetValAsp   220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 650   TATAAAGGTGTAGTTAAATACACAGTGGCAGTGAAGAAGCTTGCACAAATGGTTGAC   709

Qy 221   IleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysValTrpAlaLysCys   240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 710   ATTACTACTGAAGAACTGAAACAGCAGTTTGATCAAGAAATAAAAGTGTGGGCAAAGTGT   769

Qy 241   GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu   260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 770   CAAACATGAAAAACTTAGTAGAACTACTTGGTTTCTCAAGTGATGGAGATGACTTGTCTTA   829

Qy 361   ArgGlyGluIleThrThrLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle   380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1130   CGTGGAGAAATAACCACCAAAATCTGATATTTACAGCTTTGGTGTGGTTTTACTTGAAATA   1189

Qy 381   IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuAspLysIle   400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1190   ATAACTGGACTTCCAGCTGTCGATGAACACCGTGAACCTCAGTTATTGGATAAAATAA   1249

Qy 401   GluGluIleGluAspGluLysThrIleGluAspTyrIleAspLysLysMetAsnAsp   420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1250   GAAGAAATTGAAGATGAAAAGACAATTGAAGACTATATTGATAAAAAGATGAATGAT   1309

Qy 421   AlaAspSerThrSerValGluAlaMetTyrSerGlyAlaSerGlnCysArgHisGluLys   440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1310   GCTGATTCCACTTCAGTTGAAGCTATGTACTCTGGGGCATGCCAATGTCGGCATGAAAAG   1369

Qy 441   LysAsnLysSerProAspIleLysLysValHisGlnLeuLeuGlnGluMetThrAlaSer   460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1370   AAAAATAAGAGCCCAGACATTAAGAAAGGTTCACCAGCTGCTACAAGAGATGACAGCTTCT   1429

Qy 301   PheLeuHisGluAsnHisHisIleArgArgAspIleLysSerAlaAsnIleLeuLeuAsp   320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1070   TTTCTACATGAAAATCATCATATTCGTAGAGATATTAAAAGTGCAAATATCTTACTTGAT   1129

Qy 321   GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluArgPheAla   340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1010   GAAGCTTTTACTGCTAAAAATATCTGACTTTGGCCTTGCACGGGCTTCTGAGAAGTTTGCC   1069

Qy 341   GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu   360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 950   TTTCTACATGAAAATCATCATATTCGTAGAGATATTAAAAGTGCAAATATCTTACTTGGAT   1009

```
RESULT 8
AX196262

LOCUS       AX196262               1542 bp    DNA     linear   PAT 28-AUG-2001
DEFINITION  Sequence 4 from Patent WO0151641.
ACCESSION   AX196262
VERSION     AX196262.1  GI:15386464
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Wesche,H. and Li,S.
  TITLE     Irak-4: compositions and methods of use
  JOURNAL   Patent: WO 0151641-A 4 19-JUL-2001;
            Tularik Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1542
                     /organism="Mus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10095"
                     /note="murine IL-1 receptor-associated kinase 4 (IRAK-4)
                     cDNA"
     CDS             163..1542
                     /codon_start=1
                     /note="murine IRAK-4"
                     /protein_id="CAC60091.1"
                     /db_xref="GI:15386465"
                     /translation="MNKPLIFPSTYIRNLNVGILRKLSDFIDPQGWKKLAVAIKPSG
                     DDRYNQFHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLLVQIELFAPATLLLP
                     DAVPQTVKSLPPRAATVAQTHGPCQEKDRTSVMPKLHSHCEPPDSSSPDNRSVES
                     SDTRFHSFSFHELKSITNNPDEQPASAGGNRMGEGGFGVVYKGCVNNTIVAVKKLGAM"
```

```
BASE COUNT      421 a    392 c    423 g    306 t
ORIGIN
```

```
Alignment Scores:
Pred. No.: 1
Score:                 6.82e-166
Score:                 2017.00
Percent Similarity:    90.85%
Best Local Similarity: 83.88%
Query Match:           84.53%
DB:
```

```
US-10-001-254-16 (1-460) x AX196262 (1-1542)

Length:      1542
Matches:      385
Conservative:  32
Mismatches:    42
Indels:         0
Gaps:           0
```

```
Qy    1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuLeuAsnValGlyLeuIleArg  20
         ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db  163  ATGAACAAGCCGTTGACACCATCGACTTACATACGGTGTCTCCTTAATGTGGGAATCCTTAGG 222

Qy   21  LysLeuSerAspPheIleAspProGlnGlyTrpLysLysLeuAlaValAlaIleLys
         |||||||||||||||||||||||||||||||||||||||||||||||||:||
Db  223  AAGCTGTCCGATTTTATTGATCCTCAAGGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAA 282

Qy   41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  283  AAGCCGTCCGGCGACGACAGATACAATCAGTTCCATATAAGGAGGTTCGAAGCCTTACTT 342

Qy   61  GlnThrGlyLysSerProThrCysGluLeuLeuPheAspTrpGlyThrThrAsnCysThr  80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  343  CAGACCGGGAAGAGCCCCACGTGTGAACTGCTGTTTGACTGGGGCACACAGACTGCACA 402

Qy   81  ValGlyAspLeuValAspLeuLeuValGlnIleGluLeuPheAlaProAlaThrLeuLeu  100
         ||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db  403  GTTGGGGACCTTGTGGACCTACTGGTCCAGATTGAGCTGTTTGCCCCGGCGACTCTCCTG 462

Qy  101  LeuProAspAlaValProGlnThrValLysSerLeuProProArgAlaAlaThrVal  120
         ||||||||:||||||||||||||||||||||||:::||||||||
Db  463  CTGCCCGGATGCCGTTCCCCAAACCGTCAAAAGCCTGCCTCCTCTGAGAAGCCGAACAGTG 522

Qy  121  AlaGlnThrHisGlyProCysGlnGluLysAspArgThrSerValMetProLysLeuHisAsn  140
         :::|||||||||||||||||||||||||||||||||||||||:::
Db  523  GCACAAACACACGGCCCTTGTCAGGAAAAGGACAGGACATCCGTAACTGCCTATGCCGGAG 582

Qy  141  LeuGluGlnSerTyrMetProProAsnSerSerProGluAsnLysSerGluGluVal  160
         ||||::||||||||||||||||||||||||||:::||||||||
Db  583  CTAGACACAGCTGCGAGGCCACCGGACTCTCCAAGCCCGACAACAGAAGTGAGAGTCC 642

Qy  161  SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe  180
         |||||||||||||||||:::||||||||||||||||||||||||||||
Db  643  AGGGACACTCGGTTCCACAGCTTTCCGTTCCACGAAGCTGAAGAATCACAAACAACTTC 702

Qy  181  AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal  200
         ||||||||||||||||||||||:||:||||||||||||||||||||||||
Db  703  GACGAGCAACCCGCGTCTGCCGGTGGCAACCGGATGGGGGAGGGGGATTTGGAGTGGTG 762

Qy  201  TyrLysGlyTyrValAsnAsnThrIleValAlaValLysLysLeuGlyAlaMetValAsp  220
         |||||||||||||||||||||:::|||||||||||||||||||||:::
Db  763  TACAAGGGCTGTGTGAACAACACATCGGTGGCCGTGAAGAAGCTCGGAGCGATGGTTGAA 822

Qy  221  IleThrThrGluGluLeuLysGlnIleLeuPheAspGlnGluLysLysValMetAlaLysCys  240
         |||:::||||||||||:||||||||||||||||||||||||||||||
Db  823  ATCAGTACTGAAGAACTAAAGCAACAGTTTGATCAAGAAAATAAAGTAATGGCAAAGTGT 882

Qy  241  GlnHisGluAsnLeuValGlyLeuLeuGlyTyrCysSerAspGlyAspGluArgLeuCysLeu  260
         ||||||:::||||||||||||||||||||||||||||||||||||||||
Db  883  CACCACGAGAACCTGGTGGAGCTGCTCGGGTACTGCAGCGACGGAGACGAACTGTGCTTA 942

Qy  261  ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr  280
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  943  GTGTATGTCTTACATGCCCAACGGGTCCTTGCTGGACAGACTGTCCTGGACGATGGTACA 1002
```

```
Qy  261  ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr  280
Qy  281  ProProLeuSerTrpHisThrArgCysLysIleAlaGlnIleAlaAlaAsnGlyIleAsn  300
Qy  301  PheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp  320
Qy  321  GluAlaAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla  340
Qy  341  GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu  360
Qy  361  ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle  380
Qy  381  IleThrGlyLeuProAlaValAspProGlnLeuLeuProGlyIleLeuLeuLeuLeuAsp  400
Qy  401  GluGluLeuGluAspAspGluGluLysTyrThrLeuGluGluLeuLeuAspProGlu  420
Qy  421  AlaAspSerThrSerValGlnAlaMetTyrSerValAlaAlaSerGlnLeuHisGluLys  440
Qy  441  LysAsnArgLysProAspIleLysLysValGlnGlnLeuLeuGluMetThrAlaSer  460
```

VEISTEELKQQPEDQEIKVMATCQHENLVELGFSSSDSDNLCLVYAYMPNGSLLDRLSC
LDGTPPLSWHTRCKVAQGTANGIRFLHENHHIHRDIKSANILLDKDFTAKISDFGLAR
ASARLAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLELITGLAAVDENRE
PQLLLDIKEEIEDEEKYIEDYTDEKMSDADPASVEAMYSAASQCLHEKKNRPDIAKV
QOLLQEMSA"

```
LOCUS       AF445803                1542 bp    mRNA    linear   ROD 20-APR-2002
DEFINITION  Mus musculus interleukin-1 receptor associated kinase 4 (Irak4)
            mRNA, complete cds.
ACCESSION   AF445803
VERSION     AF445803.1  GI:20219011
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1542)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE     IRAK-4: a novel member of the IRAK family with the properties of an
            IRAK-kinase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (8), 5567-5572 (2002)
   MEDLINE  21957277
    PUBMED  11960013
REFERENCE   2  (bases 1 to 1542)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) Biology I, Tularik Inc., 2 Corporate Drive,
            South San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1..1542
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
     gene            1..1542
                     /gene="Irak4"
     CDS             163..1542
                     /gene="Irak4"
                     /note="protein kinase"
                     /codon_start=1
                     /product="interleukin-1 receptor associated kinase 4"
                     /protein_id="AAM15773.1"
                     /db_xref="GI:20219012"
                     /translation="MNKPITPSTYIRNLNVGILRKLSDFIDPQEGWKKLAVAIKKPSG
                     DRYNQFHIRRFEALLQTGKSPTELLFDWGTTNCTVGDLVDLLVQIEIAPATLLLP
                     DAVPQTVKSLLPREAATVAQTHGPCQEKDRTSVMPMPKLEHSCEPPDSSSPDNRSVES
                     SDTRFHSFSPHELKSITNNFDEQPASAGNRMGEPGFGVVYKGCVNNTIVAVKLLGAM
                     VEISTEELKQQFDQEIKWVNATCQHENLVELLLGFSSDSDNLCLVYAYMPNGSLLDRLSC
                     LDGTPPLSWHTRCKVAQGTANGIRFLHENHHIIHRDIKSANILLDKDFTAKISDFGLAR
                     ASARLAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEITGLAAVDENRE
                     PQLLLDKISEEIEDEEKTIEDYTDEKMSDADPASVEAMYSAAASQCLHEKNRRPDIAKV
                     QQILLQEMSA"
BASE COUNT       421 a    392 c    423 g    306 t
ORIGIN
```

```
RESULT 9
AF445803

Alignment Scores:
Pred. No.:
Score:                   2017.00
Percent Similarity:      90.85%
Best Local Similarity:   83.88%
Query Match:             84.53%
DB:
```

```
Qy   281  ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyIleAlaAsnGlyIleAsn  300
          ||||||||||||||||||||||||||||||::||||||||||||||||||||||||||||
Db  1003  CCACCGCTTTCCTGGCACATGAGGTGCAAGGTTGCTCAGGGGACACGGCATCAGG  1062

Qy   301  PheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp  320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1063  TTTCTGCATGAAAATCATCACATTCATAGAGATATTAAAAGTGCAAATATCTTACTAGAC  1122

Qy   321  GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGlyArgPheAla  340
          :||                                               :::||||||
Db  1123  AAAGACTTTACTGCCAAAATATCTGACTTTGGGGCTTCGGCAAGGCTAGCG  1182

Qy   341  GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1183  CAGACGGTCATGACCAGCCGAATCGTGGGCACAACGGCTTACATGGCACCCGAAGCTTTG  1242

Qy   361  ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle  380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||:||:::
Db  1243  CGGGGAGAAATAACACCCAAAATCTGACATCTACAGCTTCGGCGTGGTTCTGTTGGAGCTG  1302

Qy   381  IleThrGlyLeuProAlaValAlaAspGluHisArgGluProGlnLeuLeuLeuAspLys  400
          ||||||||||||::||||||:|||||:|||||||||:||||||||||||||||||||:||
Db  1303  ATAACCGGGCTGCCGGCTGTGGCTGACGAACCCTGAACTACTGCTGGATAATTAAA  1362

Qy   401  GluGluIleGluAspGluGluLysThrIleGluAspTyrIleAspLysLysLysMetAsnAsp  420
          ||||||||::|||||||||||||||:|::|||||::::||||||||||||||||::||
Db  1363  GAAGAGATTGAAGATGAAGAGAAGACGATTGAAGACTACATAGACAAGAAGAAGATGAAC  1422

Qy   421  AlaAspSerThrSerValGlyValAlaMetTyrSerValAlaAlaSerGlnCysLeuHisGluLys  440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1423  GCGGGACCCTGCTTCGGTGGAGCAATGTACTCTGTCGCCGCCAGTCAGTGCCTGCATGAGAAG  1482

Qy   441  LysAsnLysArgProAspIleLysLysValGlnGlnIleLeuLeuGlnGluMetThrAla  459
          ||||||||||:|||||||||||||||||||||||::|||::|||||:::||
Db  1483  AAAAACAGAACGGCGCCAGACATTGCAAAAGGTTCAACAGCTGCTACAAGAGATGTCTGCT  1539
```

```
US-10-001-254-16 (1-460) x AF445803 (1-1542)

Alignment Scores:
Pred. No.:
Score:                   6.82e-166
Percent Similarity:      90.85%
Best Local Similarity:   83.88%
Query Match:             84.53%
DB:                      10

Length:        1542
Matches:       385
Conservative:  32
Mismatches:    42
Indels:        0
Gaps:          0
```

```
Qy    1  MetAsnLysProIleThrProSerThrTyrValArgAsnLeuAsnValGlyIleLeuArg  20
         |||||:|||:||||||||||||||||||||||||||||||||:::||||||||||
Db  163  ATGAACAAGCCGTTACCATCGAACACATACGACAGAAACTTAAATGTGGGGATCCTTAGG  222

Qy   21  LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
         ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db  223  AAGCTGTCGGATTTTATTGATCCTCAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAA  282

Qy   41  LysProSerGlyAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu  60
         ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db  283  AAGCCGTCCGGCGACAGATACAATCAGTTCCATATAAGGAGGTTCGAAGCCTTACTT  342

Qy   61  GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr  80
         |||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db  343  CAGACCGGGAAGAGCCCCACCTCTGAACTGCTGTTTGACTGGGGAACTACAAACTGCACA  402

Qy   81  ValGlyAspLeuValAspLeuLeuValGlnIleGluIleAlaProAlaThrLeuLeu  100
         ||||||||||:|::||||||||||||||||||||||||||||||||:||::
Db  403  GTTGGCGACCTTGTGGATCTACTGGTCCAGATTGAGTCTGTTGCCCCGGCCACTCTCCTG  462

Qy  101  LeuProAspAlaValProGlnThrValAlaAsnThrLeuAlaAsnThrLeuAlaAspThrVal  120
         |||||||||:::|||||:||||||||||:|||||||||:||
Db  463  CTGCCCGGATGCCGGTTCCCCAAACCGTCAAAAAGCCTCGGACACAGTG  522

Qy  121  ValGlyAspLeuValAlaAspLeuLeuLeuGlyPheSerSerAspSerAspAsnLeuCys  140
         :::||||||||:|||::|||||||||||||||||:::||||||||||||||||||||||:
Db  523  GCACAAACACAGCGCCCTTGTCAGGAAAAGGACAGACATCCGTAAAGCCTAATGCCGAAG  582

Qy  141  LeuGluGlnSerIleTyrMetProAsnProAspSerSerSerProGluAsnLeuSerLeuGluVal  160
         ||||||||:||||||||||||||||||||||||::||||:|
Db  583  CTAGACACAGCTGCGAGCCACCGGACTCCTCAAGCCCAGACAGAAGGTTAGAGTCC  642

Qy  161  SerAspThrArgPheHisSerPheSerProHisTyrGluLeuLysSerAsnValThrAsnAsnPhe  180
         ||||||||:|||||||:|||::|||||||:||
Db  643  AGCGACACATCCGGTTCCACAGCTTCTCCGTTCCGTTACCATGAAGCATCACAAAACACTTC  702

Qy  181  AspGluGlnArgProIleSerValAlaGlyAsnLysMetGlyGluGlyTyrGlyPheGlnLys  200
         |||||::||||||:|||||||:|||::||||||||||||||||||||||::|||||:||
Db  703  GACGAGCAACCCGGCGTCTGCGCGGTGGCAACCCGGAGGGGGGGAATTGGAGTGGTCGTG  762

Qy  201  TyrLysGlyTyrValAsnAsnThrIleValAlaValAlaValAlaLysMetValLeuAsp  220
         ||||||||||||||||||::::|||||||||::||||||||||::||||||||||||:||
Db  763  TACAAGGGCTGTGTGAACAACACCATCGTCGCGGTGAAGCTCCTGGGAGCCGATGGGTTGAA  822

Qy  221  IleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysLysValValMetAlaLysCys  240
         |||||:::||||||||||:|||||||:|||||||::|||||||||||||:||
Db  823  ATCAGTACTGAAGAACTAAAGCAACAGTTTGATCAAGAAATTAAAGTAATGGCCAACGTGT  882
```

```
Qy  241 GlnHisGluAsnLeuValGluLeuGlyPheSerSerAspGlyAspLeuCysLeu 260
        |||||||||||||||||||||||||||||||||:::|||
Db  883 CAGCACGAGAACCTGGTGGAGCTGGGCTTCTCCAGCGACGGACAACCTGTGCTTA 942

Qy  261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr 280
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  943 GTGTATGTTACATGCCAAACGGGTCTCCTTGGCTGGACAGACTGTCCTGTGATGGTACA 1002

Qy  281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyIleAlaAsnGlyIleAsn 300
        |||||||||||||||||||||||||||||||||||||||||||||||
Db 1003 CCACCGCTTTCCTGGCACATGAGGTGCAAGGTTGCTCAGGGCATCGCCAATGGCATCAGG 1062

Qy  301 PheLeuHisGluAsnHisIleHisArgAspIleAlaAsnIleLeuLeuAsp 320
        ||||||||||||||||||||||||||||||||||||
Db 1063 TTTCTGCATGAAAATCATCACATTCATAGAGATATTAAAAATCTTACTAGAC 1122

Qy  321 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla 340
        :::|||||||||||||||||||||||:::
Db 1123 AAAGACTTTACTGCCAAAATATCTGACTTTGGGCTTGCAAGGGCTAGCG 1182

Qy  341 GlnThrValMetThrSerArgIleValAlaGlyThrThrAlaTyrMetAlaProGluAlaLeu 360
        |||||||||||||||||||||||||||||||||
Db 1183 CAGACCGGTCATGACAAGCACCGGAATCGTGGCAGGCACCACCGCCTACATGGCACCCGAAGCTTTG 1242

Qy  361 ArgGlyGluIleThrProLysSerAspPheGlyLeuValValLeuLeuGluIle 380
        ||||||||||||||||||||||||||||||:::
Db 1243 CGGGGAGAAATAACACCCAAAATCTGACATCTACTACAGCTGTTCTTGTGGAGCTG 1302

Qy  381 IleThrGlyLeuAlaAlaAspProGluHisArgGluProGlnLeuLeuLeuAspAspLys 400
        |||||||||||:::|||||||||||||||||||||||||||||:::
Db 1303 ATAACCGGGCTGGCGGCTGATGAAAAACCTGAACCTACACTAACTGGATATTAAA 1362

Qy  401 GluGluIleGluAspGluGluLysThrIleGluAspTyrThrGluAspLysMetThrArgAla 420
        |||||||||:::|||||||||||||||||||||||||
Db 1363 GAAGAGATTGAAGATGAAGAGAAGACCATTGAAGATTACACGGAAGATAAGATGAGCGAT 1422

Qy  421 AlaAspSerThrSerValGluAlaAlaMetTyrSerValAlaAlaSerGlnCysLeuHisGluLys 440
        |||:::||||||||||||||||||||||||||||||||||||||
Db 1423 GCGGACACCCTGCTTCGGTGGTGAAGCAAATGTACTCTGCTGTAGCCGATGGAGATGTCTGCT 1482

Qy  441 LysArgAsnLysArgProAspIleLysLysValGlnLeuLeuGlnGluMetThrAla 459
        |:::||||||||||||||||||||||||||||:::|||
Db 1483 AAAAAACAGACAGCCCAGACAGATTGCCAAAGGTTCAACAGCTGCTACAAGAGATGTCTGCT 1539
```

RESULT 10
```
LOCUS       BC051676                2431 bp    mRNA    linear   ROD 10-JUN-2003
DEFINITION  Mus musculus interleukin-1 receptor-associated kinase 4, transcript
            variant 1, mRNA (cDNA clone MGC:60994 IMAGE:30017484), complete
            cds.
ACCESSION   BC051676
VERSION     BC051676.1  GI:30354527
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 2431)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE  22382257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2431)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
```

```
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Minoru Ko
            cDNA Library Preparation: Yulan Piao and Minoru Ko
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne,P.H., Garcia,A.M., Lu,X., Hulyk,S.W., Loulseged,H.,
            Kowis,C.R., Sneed,A.J., Martin,R.G., Muzny,D.M., Nanavati,
            A.N., Gibbs,R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 111 Row: i Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 23943897.
FEATURES             Location/Qualifiers
     source          1..2431
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="MGC:60994 IMAGE:30017484"
                     /tissue_type="Embryo, whole, mouse, 7.5-dpc"
                     /clone_lib="NIA Mouse 7.5-dpc Whole Embryo cDNA Library
                     (Long)"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1"
     gene            1..2431
                     /gene="Irak4"
                     /note="synonyms: NY-REN-64, 9330209D03Rik"
                     /db_xref="LocusID:266632"
     CDS             168..1547
                     /codon_start=1
                     /product="interleukin-1 receptor-associated kinase 4,
                     isoform a"
                     /protein_id="AAH51676.1"
                     /db_xref="GI:30354528"
                     /db_xref="LocusID:266632"
                     /translation="MNKPLIPFSTYIRNLNVGILRKLSDFIDPQEGWKKLAVAIKPSG
                     DRYNQFHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLLVQIELFAPATLLLP
                     DAVPQTVKSLPPREAATVAQTHGPCQEKDRTSVMPKLEHSCEPFDSSSPDNRSVES
                     SDTRFHSFSFHELKSITNNFDEQPASAGNRMGEGGFGVVYKGCVNNTIVAVKKLGAM
                     VEISTEELKQQFDQEIKVMATCQHHNIVELLGFSSDSDNLCLVYAYMPNGSLLDRLSC
                     LDGTPPLSWHTRCKVAQGTANGIRFLHENHHIHRDIKSANILLDRDFTAKISDFGLAR
                     ASARLAQTVMTSRIVGTTIAVMAPEALRGEITPKSDIYSFGVLLLEITGLAAVDENRE
                     PQLLLDIKEEIEDDEEKTIEDYTEDKMSDADPASVEAMYSAASQCLHEKKRRPDIAKV
                     QQLLQEMSA"
BASE COUNT      656 a    580 c    618 g    577 t
ORIGIN
```

Alignment Scores:
Pred. No.:

1.51e-165   Length: 2431

Score: 2016.00   Matches: 385
Percent Similarity: 90.63%   Conservative: 31
Best Local Similarity: 83.88%   Mismatches: 43
Query Match: 84.49%   Indels: 0
DB: 10   Gaps: 0

US-10-001-254-16 (1-460) x BC051676 (1-2431)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg   20
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 168 ATGAACAAGCCGTTGACCCCATCAACGACATACATACGTGGGGATCCTTAGG   227

Qy   21 LysLeuSerAspPheIleAspProGlnGlnGlyGlyTrpLysLysLeuAlaValAlaIleLys   40
         |||||||||||||||||||||||||||||||||||||||||||||
Db 228 AAGCTGTCGGAATTTTATTGATCCTCAAGAGGGTGGAAGAAAATTAGCAGTAGCTATCAAA 287

Qy   41 LysProSerGlyAspAspArgTyrArgAsnGlnPheHisIleLeuArgPheGluAlaLeuLeu   60
         ||||||||||||||||||||||||||||||||||||||
Db 288 AAGCCGTCCGGCGACGACAGATACAGAAATCAGTTTCATATAAGGAGAGATTGAAGCTCTT 347

Qy   61 GlnThrGlyLysSerProThrSerGlnLeuLeuLeuPheGlyThrThrAsnCysThr   80
         |||||||||||||||||||||||||||||
Db 348 CAGACCGGGAAGAGCCCAACTAGTCAGCTGCTGTTGCCAACCTTACTT   407

Qy   81 ValGlyAspLeuValAlaSerProIleLeuGlnAsnGluPhePheAlaProAlaSerLeuLeu   100
         ||| ||||::::||
Db 408 GTTGGCGACCTTGTTGGATCTACTACTGGTCCAGATTGAGCTGTTTGCCCCCGCCAGTCTGCTG 467

Qy 101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal   120
         |||||||:::||
Db 468 CTGCCCGGATGCCGTTCCCCAAACCGTCAAAAAGCCTCGCCTGAGAAGCGGCACACAGTG   527

Qy 121 GlnIleLysGlnMetProPheCysSerAspLysArgThrLeuMetThrProValGlnAsn   140
         :::||||||
Db 528 GCACAAACACACGGCGCCTTGTCAGGAAGACAGACATCCGTAATGCCTATGCCGAAG   587

Qy 141 LeuGluGluGlnThrTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal 160
         ||||||||||||||||
Db 588 CTAGAAGAACAGACCTACATGCCGCCAGACTCCTCAAGCCCAGAACAGAAGTGTAGAGTCC   647

Qy 161 SerAspThrArgGlyPheHisSerPheTyrGlyIleLeuLysAsnValThrAsnAsnPhe   180
         :::|||||||||:::||
Db 648 AGCGACACTCGGGTTCCACAGTTTCTCGTTCCATGAACTGAAGAACATCACAAACAACTTC 707

Qy 181 AspGluArgProIleSerGluValGlyIleGlyAsnGlyGlyValGlyGlyPheGluGlyValVal 200
         ||||:::||
Db 708 GACGAGCAACCCCGCGTTCTGCCGGTGGCACCGGATGGAGGAGGGGATTTGGAGTGGGGTG   767

Qy 201 TyrLysGlyTyrValAsnAsnThrThrValAlaValLysSerLeuAlaAlaMetValAsp   220
         ||||||||||||
Db 768 TACAAGGGCTGTGTGAACAACACCATCGTGGCGGTGAAGAGCCTCGGAGCGGATGGTTGAA   827

Qy 221 IleThrThrGlnGluLeuLysGlnIleLysGlyAlaGlnIleLeuLysLysValMetAlaLysCys 240
         |||||||||||
Db 828 ATCACGACTCAGGAACTAAAGCAAATTAAAGGTTGATCAGAAATTAAAGTGACAACGTGT   887

Qy 241 GlnHisGluAsnLeuValGluLeuGlyLeuLeuPheSerSerGluAspGlyAspLeuCysLeu   260
         ||||||||||||||
Db 888 CAGCACGAGAACCTGGTGGAGCTCGGGCTTCTCAGCGACGAACCTGTGCTTA   947

Qy 261 ValThrTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspAsp   280
         |||||||:::|||
Db 948 GTGTATGCCTTACACGCCCAACCGGTCCCTTGCTCGACAGACTGTCCTCCGATGGTACA   1007

Qy 281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyValAlaAlaAsnGlyIleAsn   300
         |||:::
Db 1008 CCACCGCCTTCCTCGGCACAAGGTGCTGCAGGGCACACAAGGGCATCAGG   1067

Qy 301 PheLeuHisGluAsnHisHisIleHisArgAspLeuLysSerAlaAsnIleLeuLeuAsp   320
         |||||||:::
Db 1068 TTTCTGCATGAAAATCACCATTCATAGAGATTTAAAAGTGCAAATATCTTACTAGAC   1127

Qy 321 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla   340
         ||||||||||||
Db 1128 AGAGACGGTTCATTGCCAAAAATATCTGACTTTGGGCTTGCACGGGCTAGGGCTAGCG   1187

Db 1128 AGAGACGTTACTTGCCAAAAATATCTGACTTTGGGCTTGCACGGGCTAGCG   1187
         ||||||||||||||||||||| :::

Qy 341 GlnThrValMetThrSerArgIleValAlaGlyThrThrAlaTyrMetAlaProGluAlaLeu   360
         |||||||||||||||||||||||
Db 1248 CGGGGAGCAGAAAATAACACCCAAAATCTGACTTTGGGCTTGCACGGGCTAGGGCTAGGCTG   1307

Qy 361 ArgGlyGlyIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle   380
         |||||||||||||||||||||||
Db 1368 GAAGAGATTGAAGACCGGGTGGAAGAGAAGCCGATTGAAAAGCTGATTAAAGGAGCGGAT   1427

Qy 381 IleThrGlyLeuProAlaValAlaAspGluHisArgGluProGlnLeuLeuLeuAsnIleLys   400
         |||||||||||||||||||||||||||||
Db 1308 ATAACCGGGCTGGGCGGCCTGTCGATGAAAAACCGTGAAACCTCAACTACTGCTGATAATTAAA 1367

Qy 401 GluLeuIleGluGluAspGluGlyLysThrIleGluGluAspTyrIleLeuLysMetAsnAsp   420
         |||||||||||||||::::
Db 1368 GAAGAGATTGAAGACCGGGTGGAAGAGAAGCCGATTGAAAAGCTGATTAAAGGAGCGGAT   1427

Qy 421 AlaAspSerThrSerGlnIleAlaMetTyrSerValAlaAlaSerGlnAsnCysLeuHisLys   440
         ||||||||||:::|||
Db 1428 GCGGACGCCCTTCGGTTGGAAGCCAATGTACTCGTGCTAGCCAGTTGCTCTCGGCCCATGAGAAG 1487

Qy 441 LysAsnLysArgProAspSerIleLysLysValGlnGlnIleLeuGlnGlnIleGluMetThrAla 459
         ||||:::||||||||:::|||
Db 1488 AAAAACAACGGCCCAGACATTGCCAAAGGTTCAACAACTGCTACAAGAGATAGTCTGCT   1544

RESULT 11

LOCUS   BD155790   2213 bp   DNA   linear   PAT 17-JAN-2003
DEFINITION   Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD155790
VERSION   BD155790.1 GI:27861548
KEYWORDS   JP 200219363-A/10633.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2213)
AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
             Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE   Primer for synthesizing full-length cDNA and use thereof
JOURNAL   Patent: JP 200219363-A 10633 09-JUL-2002;
             HELIX RESEARCH INSTITUTE
COMMENT   OS   Homo sapiens (human)
             PN   JP 200219363-A/10633
             PD   09-JUL-2002
             PF   28-JUL-2000 JP 2000280990
             PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
             PI   SAITO,
             PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
             PI   KEIICHI NAGAI,TETSUJI OTSUKI
             PC   C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
             10,
             PC   C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/00,C12N5/ PC
             PC   C12P21/02,C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
             Primer for synthesizing full-length cDNA and use thereof FH Key
             Location/Qualifiers
FEATURES   Location/Qualifiers
     source   1..2213
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     FT   CDS   (977)..(1864).
                     Primer for synthesizing full-length cDNA and use thereof FH Key
     CDS   Location/Qualifiers
BASE COUNT   642 a   488 c   516 g   567 t
ORIGIN

Alignment Scores:
Pred. No.:   1. .2213
Score: 9.7e-146   Length: 2213
Percent Similarity: 87.34%   Matches: 352
Best Local Similarity: 87.34%   Conservative: 0
Query Match: 74.92%   Mismatches: 1
                                         Indels: 50
                                         Length: 1787.50

US-10-001-254-16.rge

DB: US-10-001-254-16 (1-460) x BD155790 (1-2213)   6   Gaps: 1

```
Qy      1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuLeuArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    751 ATGAACAAAACCCATAACACCATCAACAACATATGTCGGTGCCTGCCTCAATGTTGGACTATTAGG 810

Qy     21 LysLeuSerAspPheIleAspProGlnGlyTrpLysGlyLeuValAlaIleIleLys 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    811 AAGCTGTCAGATTTTATTGATCCTCAAGGAAGGAAGGGATTAGCTGTAGCTATTAAA 870

Qy     41 LysProSerGlyAspArgTyrAsnGlnPheHisIleArgArgPheGlnIleLeuLeu 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    871 AAACCATCTGGTGATAGATACAATCAGTTTCACATAAGA------------------ 912

Qy     61 GlnThrGlyGlySerProThrSerGlyGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
          ------------------------------------------------------------
Db    912 ------------------------------------------------------------ 912

Qy     81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeu 100
          ------------------------------------------------------------
Db    912 ------------------------------------------------------------ 912

Qy    101 LeuProAspAla-ValProLysThrThrLeuProSerTyrGluAlaIleThrVa 120
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    913 ------------TGCTGTGTCCCAAAACTGCTAATACACTRACCTTCTAAAGAGAGAGCTAAAACAGT 963

Qy    121 lGlnGlnLysGlnMetProPheCysAlaAspLysArgAspArgThrLeuMetThrProValGlnAs 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    964 TCAGCAAAAACAGATGCCTTTCGTGTGGTAATAAAAGGGGAGGGAGGGATTTTGGACTTGT 1023

Qy    141 nLeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVa 160
          --------------------------------------------------------------
Db   1024 TCTTGAACAAAGCTAATAGCCACCTGACTCCTTCAAGTCCCAGAAAAATAAAAGTTTAGAAGT 1083

Qy    161 lSerAsnThrArgPheHisSerPheSerPheTyrGluLeuLysAspAsnSerValThrAsnAsnPh 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1084 TAGTGGATACACGTTTTCACAGTTTTCATTTTATGAATTGGAGAAGTGTCACAAATAACTT 1143

Qy    181 eAspGluArgProIleSerSerValGlyGlyAsnLysMetGlyGlyPheGlyGlyValVa 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1144 TGATGAACGACCCATTTCTGTGGTGGTAATAAAATGGGAGGGAGGGATTTGGACTTGT 1203

Qy    201 lTyrLysGlyTyrValAsnAsnThrThrValAlaAlaValLysSerLeuAlaAlaMetValAs 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1204 ATATAAAAGGCTACGTAAATAACACACAGTGCGCAGTGAAGAAGCTTGCAGCAATGGTTGA 1263

Qy    221 pIleThrThrGluGlnLeuGlnLeuIleGlnAsnPheAspGlnGluLeuLysValMetAlaLysSerCy 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1264 CATTACTACTGAAGAACTGAAACAGCAGTTTGATCAAGAAATAAAAGTAATGGCAAAGTG 1323

Qy    241 sGlnHisGlnLeuAsnValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLe 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1324 TCAACATGAAAACTTAGTAGAACTACTGGGTTTCTCAAGTGATGGAGATGACCTGTGCTT 1383

Qy    261 uValThrTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyTh 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1384 AGTAACATATGTTTACATGCCTAACGGTTCATCTGCTAGACAGACTCTCTTGCTTGGATGGTAC 1443

Qy    281 rProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAs 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1444 TCCACCACTTTCTTGCCACATGAGATGCAAGATTGCTCAGGGTGCAGCTAATGGCATCAA 1503

Qy    301 nPheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAs 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1504 TTTTCTACATGAAAAACATCATATTCATAGAGATATTAAAAGTGCAAATAACTTACTTGGA 1563

Qy    321 pGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGlyLysPheAl 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1564 TGAAGCTTTTACTGCTAAAAATATCTGACTTTGGCCTTGCACGGGCTTCGTGGAAAGTTTGC 1623

Qy    341 aGlnThrValMetThrSerArgIleValValGlyThrThrAlaTyrMetAlaProGluAlaLe 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1624 CCAGACAGTCATGACTAGCAGAATTGTGGGACAACAGCTTATATGGACCAGAAGCTTT 1683

Qy    361 uArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIl 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1684 GCGTGGAGAAATAACACCCAAAATCTGATATTTACAGCTTTGGTGTGGTTTTACTAGAAAT 1743

Qy    381 eIleThrGlyLeuProAlaValAspProGlyHisArgGluProGlnLeuLeuLeuAspIlely 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1744 AATAACTGGACTTCCAGCTGTGGATCCAGGACACCGTGAACACCCTCAGTTATTGCTAGATATTAA 1803

Qy    400 sGluGlu 402
          |||||||
Db   1804 AGAAGAA 1810
```

RESULT 12

AK027301

LOCUS   AK027301   2213 bp   mRNA   linear   PRI 01-AUG-2002

DEFINITION   Homo sapiens cDNA FLJ14395 fis, clone HEMBA1003250, weakly similar
             to PROTEIN KINASE APKIA (EC 2.7.1.-).

ACCESSION   AK027301

VERSION   AK027301.1 GI:14041890

KEYWORDS   oligo capping; fis (full insert sequence).

SOURCE   Homo sapiens (human)

  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1

  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
             Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
             Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
             Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
             Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
             Ishii,S., Kawai,Y., Saito,K., Yanamoto,J., Wakamatsu,A.,
             Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.

  TITLE   NEDO human cDNA sequencing project

  JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 2213)

  AUTHORS   Isogai,T. and Otsuki,T.

  TITLE   Direct Submission

  JOURNAL   Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
             Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
             (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT   NEDO human cDNA sequencing project supported by Ministry of
             Economy, Trade and Industry of Japan; cDNA full insert sequencing;
             Research Association for Biotechnology; cDNA library construction,
             5'- & 3'-end one pass sequencing and clone selection: Helix
             Research Institute (supported by Japan Key Technology Center etc.)
             and Department of Virology, Institute of Medical Science,
             University of Tokyo.

FEATURES   Location/Qualifiers
     source   1..2213
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1003250"
                     /tissue_type="whole embryo, mainly head"
                     /clone_lib="HEMBA1"
                     /dev_stage="embryo, 10 weeks"
                     /note="cloning vector: pME18SFl3"

BASE COUNT   642 a   488 c   516 g   567 t

ORIGIN

Alignment Scores:
Pred. No.:
Score: 9.7e-146   Length: 2213
Percent Similarity: 1787.50   Matches: 352
Best Local Similarity: 87.34%   Conservative: 0
Query Match: 87.34%   Mismatches: 1
DB: 74.92%   Indels: 50
                                            9   Gaps: 1

US-10-001-254-16 (1-460) x AK027301 (1-2213)

```
Qy      1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
Db    751 ATGAACAAACCCATAACACCATCAACACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1684 GCGTGGAGAAATAACACCCAAATCTGATATTTACAGCTTGGTGTGGTTTTACTAATTAAT 1743

Qy     21 LysLeuSerAspPheIleAspProGlnGlyLysThrProLysLysLeuValAlaValAlaIleLeuLys 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 LysLeuSerAspPheIleAspProGlnGlyLysThrProLysLysLeuValAlaValAlaIleLeuLys 40
Qy    912 ------ 912
Db    811 AAGCTGTCAGATTTTATTGATCCTCAAGGAAAGGATGGAAGAAAGTTAGCTGTAGCTATTAAA 870
Db    912 eileThrGlyLeuProAlaValAspGluHisArgGluHisArgGlnLeuLeuAspLeuy 400

Qy     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleLeuArgArgPheGluAlaLeuLeu 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleLeuArgArgPheGluAlaLeuLeu 60
Db    871 AAACCATCTGGTGATGAAAGATACAAATCAGTTTCACATAAGA------ 912
Db   1744 AATAACTGGACTTCCAGCTGTGGATGAACACCGTGAAACCTCAGTTATTGCTAGATATTAA 1803

Qy     61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
Db    912 ------ 912
Db    913 ---TGCTGTTCCCAAAACTGCTAATACACTACCTTCTAAAGAACGCTAATAACAGT 963

Qy     81 ValGlyAspLeuValAspLeuLeuIleGluIleAsnGluPhePheAlaProAlaSerLeuLeu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 ValGlyAspLeuValAspLeuLeuIleGluIleAsnGluPhePheAlaProAlaSerLeuLeu 100
Db    912 ------ 912

Qy    101 LeuProAspAla-ValProLysThrAlaAsnThrLeuProSerIleTyrGluValThrVa 120
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 LeuProAspAla-ValProLysThrAlaAsnThrLeuProSerIleTyrGluValThrVa 120
Db    964 TCAGGACAAAACAGATGCCTTTCTGTGACAAAAGACACAAGACACCTGTGCAGAA 1023

Qy    121 IleGlnIleLysGlnMetProPheCysAspAspLeuTyrGlyLeuGlyValGlnAsn 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 IleGlnIleLysGlnMetProPheCysAspAspLeuTyrGlyLeuGlyValGlnAsn 140
Db   1024 TCTTGAACAAAAGCTATATGCCACTTGACTCCTCAAGTCCCAAAAATAAAAAGTTTAGAAGT 1083

Qy    141 eAspGluArgProIleSerValGlyGlyAsnLysMetGlyGlyValVa 160
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    141 eAspGluArgProIleSerValGlyGlyAsnLysMetGlyGlyValVa 160
Db   1144 TGATGAACGACACCCATTTCGTTGGTTGGTAATAAAATGGGAGGAGGAGGAGTTGGT 1203

Qy    161 nLeuGluGlnSerIleTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 nLeuGluGlnSerIleTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVa 180
Db   1084 TAGTGAATACACGTTTTCACAGTTTTTCATTTTATGAATTGAAGAATGTCACGAAAATAACTT 1143

Qy    181 lSerAspThrArgPheHisSerPheSerPheTyrGlyLeuLeuLysLeuAlaIleMetValAs 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 lSerAspThrArgPheHisSerPheSerPheTyrGlyLeuLeuLysLeuAlaIleMetValAs 200
Db   1204 ATATAAAGGCTACGTTAAATAACAACTGTGAAGAAGACTTGCACGATTGCACCAATGTTGA 1263

Qy    201 lTyrLysGlyTyrValAsnAsnThrThrValAlaLysValGlyLeuAlaLeuAlaMetValAs 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 lTyrLysGlyTyrValAsnAsnThrThrValAlaLysValGlyLeuAlaLeuAlaMetValAs 220
Db   1264 CATTACTACGAAACTGAAGAACACAGCAGTTTGATCAGAAAATAAAAGTAATGGCAAAGTG 1323

Qy    221 pIleThrThrGluGluLeuValGlnGlnProAspAspLysAlaAsnAlaAlaAsnIleLeuLeuAs 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 pIleThrThrGluGluLeuValGlnGlnProAspAspLysAlaAsnAlaAlaAsnIleLeuLeuAs 240
Db   1324 TCAACATGAAAAACTTAGTAGAACTACTTGGTTTCTCAAGTGAGTAGATGACCTCTGCTT 1383

Qy    241 sGlnHisGluAsnLeuValGlyLeuLeuGlyPheSerSerAspAspLeuCysLe 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 sGlnHisGluAsnLeuValGlyLeuLeuGlyPheSerSerAspAspLeuCysLe 260
Db   1384 AGTATATANGTTTACATGCCATAAACGGTTCAATTGCTAGACACTCTCTTGCTTGATGGGTAC 1443

Qy    261 uValThrTyrValTyrMetProAsnGlySerLeuAspArgLeuSerCysLeuAspGlyTh 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 uValThrTyrValTyrMetProAsnGlySerLeuAspArgLeuSerCysLeuAspGlyTh 280
Db   1504 TTTTCTACATGAAAAATCATCATATTTCACGATAATAAAAGTGCAAAATATCTTACTGGA 1563

Qy    281 rProProLeuSerTrpHisMetArgGlyLysIleAlaGlnIleAlaLeuAsnGlyIleAs 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 rProProLeuSerTrpHisMetArgGlyLysIleAlaGlnIleAlaLeuAsnGlyIleAs 300
Db   1444 TCCACCACTTTCCTTGGCACATGAAAGGATGCAAGATTGCTCAGGGTGCGAGCTAATGGCATCAA 1503

Qy    301 nPheLeuHisGluAsnHisHisIleHisSerArgAspIleLeuGluLysSerAlaAlaAsnIleLeuLeuAs 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 nPheLeuHisGluAsnHisHisIleHisSerArgAspIleLeuGluLysSerAlaAlaAsnIleLeuLeuAs 320
Db   1564 TGAAGCTTTTACTCTGCTAAAATATCTGACTTTGGCCTTGCACGGGCCTTCTGAGAAGTTTGC 1623

Qy    321 pGluAlaPheThrAlaLysIleLeuSerGluAspIleLeuGluLeuAlaArgLeuAlaSerGlyLysPheAl 340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 pGluAlaPheThrAlaLysIleLeuSerGluAspIleLeuGluLeuAlaArgLeuAlaSerGlyLysPheAl 340
Db   1624 CCAGACAGTCATGACTAGCAGAATTGTGGACAACAGCTTTATATGGCACCAGAAGCTTT 1683

Qy    341 aGlnThrArgValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLe 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341 aGlnThrArgValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLe 360
Db   1684 AGAAGAA 1810

Qy    361 uArgGlyGluIleThrProLysSerAspIleThrTyrSerPheGlyValValLeuLeuGluIl 380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 uArgGlyGluIleThrProLysSerAspIleThrTyrSerPheGlyValValLeuLeuGluIl 380
```

US-10-001-254-16 (1-460) x AR223870 (1-833)

Alignment Scores:
Pred. No.:
Score: 1169.00   Length: 833
Percent Similarity: 93.21%   Matches: 246
Percent Similarity: 92.83%   Conservative: 1
Best Local Similarity: 48.99%   Mismatches: 8
Query Match: 6   Indels: 10
DB: 1.56e-92   Gaps: 3

```
Qy      1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
Db    250 ATGAACAAACCCATAACACCATCAACACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 109

Qy     21 LysLeuSerAspPheIleAspProGlnGlyLysThrProLysLysLeuValAlaValAlaIleLeuLys 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 LysLeuSerAspPheIleAspProGlnGlyLysThrProLysLysLeuValAlaValAlaIleLeuLys 40
Db    110 AAGCTGTCAGATTTTATTGATCCTCAAGGAAAGGATGGAAGAAAGTTAGCTGTAGCTATTAAA 169

Qy     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleLeuArgArgPheGluAlaLeuLeu 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleLeuArgArgPheGluAlaLeuLeu 60
Db    170 AAACCATCTGGTGATGAAAGATACAAATCAGTTTCACATAAGAACGTTTCAATTGCACA 229

Qy     61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
Db    230 CAAACTGGAAAAAGTCCCACTTCGAATTACTGTTTTGACTGGGGCCACCACAAAATGCACA 289

Qy     81 ValGlyAspLeuValAspLeuLeuIleGluIleAsnGluPhePheAlaProAlaSerLeuLeu 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 ValGlyAspLeuValAspLeuLeuIleGluIleAsnGluPhePheAlaProAlaSerLeuLeu 100
Db    290 GTTGGTGATCTTGTGGATCTTTTGATCGAAATAAATGAATTTTTTGCTCCTGCGAGTCTTTTG 349

Qy    101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerIleTyrGluValThrVal 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerIleTyrGluValThrVal 120
Db    350 CTCCCAGATGCTGTTCCCAAAACTGCTAATACACTACCTTCTAAAGAACGCTAATAACAGT 409

Qy    121 GlnIleLysGlnMetProPheCysAspAspLeuTyrGlyLeuGlyValGlnAsn 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GlnIleLysGlnMetProPheCysAspAspLeuTyrGlyLeuGlyValGlnAsn 140
Db    410 CAGCAAAAACAGATGCCTTTCTGTGACAAAAGACACAAGACACCTGTGCAGAAT 469

Qy    141 LeuGluGlnSerIleTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 LeuGluGlnSerIleTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal 160
Db    470 CTTGAACAAAAGCTATATGCCACTTGACTCCTCAAGTCCCAAAAATAAAAAGTTTAGAAGT 529
```

RESULT 14

BC045381

```
LOCUS       BC045381                1719 bp    mRNA    linear   VRT 10-FEB-2003
DEFINITION  Danio rerio, Similar to interleukin-1 receptor-associated kinase 4,
            clone MGC:55553 IMAGE:2642773, mRNA, complete cds.
ACCESSION   BC045381
VERSION     BC045381.1  GI:28278875
KEYWORDS    MGC.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 1719)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
            cDNA Library Preparation: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 101 Row: m Column: 22
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..1719
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="MGC:55553 IMAGE:2642773"
                     /tissue_type="Whole body, adult, (one male and one female,
                     including unfertilized eggs)"
                     /clone_lib="Sugano Kawakami_zebrafish_DRA"
                     /lab_host="DH10B"
     CDS             122..1270
                     /note="Vector: pME18S-FL3"
                     /codon_start=1
                     /product="Similar to interleukin-1 receptor-associated
                     kinase 4"
                     /protein_id="AAH45381.1"
                     /db_xref="GI:28278876"
                     /translation="MSDVTNPVRKLRYSALRALADLLDPDTWRSIMADISRPCGE
                     PRYTQMHMRRFEACVLCQKSPTMELLFDWGTDSDCTVGDLVEILIIRHQLFAAVTLLPD
                     HSVCHAHTGSWWCEERAPASAVCLQACEITQTVEDNSNKPQKISKVEDDSNKPVQEQ
                     FRGRWGDKHVAVKKLNPLDGSSYEDLRKQFNQEIQTLRSLSHENVLRVLGCSGPPL
                     CVVFELMVNGSLLERLACAERTPALTWRNRCWITVGAANGLSYLHTHAHIHRDVKSAN
                     ILLDEGFVAKISDFGLTRSAAAGAGGFVWASSSGRESGSCTVAGDEG"
BASE COUNT      492 a    399 c    477 g    351 t
ORIGIN
```

Alignment Scores:
```
Pred. No.:              8.82e-76          Length:        1719
Score:                  981.00            Matches:        199
Percent Similarity:     61.34%            Conservative:    93
Best Local Similarity:   4.81%            Mismatches:     121
Query Match:            41.11%            Indels:          63
DB:                                       Gaps:             4
```

US-10-001-254-16 (1-460) x BC045381 (1-1719)

```
Qy    5 IleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSerAsp 24
        ::|:|||::|||||||||||||:||||:||::|::|:|:::|||
Db  131 GTCACGCCGGAACACACCGGTTCGAAAGTCCGGTACTCCGGGCTTAAGAGCTTTAGCAGAC 190

Qy   25 PheIleAspProGlnGluGlyTrpLysLysLeuAlaIleLeuLysLysProSerGly 44
        ::||||||||:|:||||::|
Db  191 CTGCTGCAGACCCCGCAGGAGCGACACCTGGAGGAGCAATCATGCGGACAATCAGCAGACCCTGCGGG 250

Qy   45 AspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLys 64
        :::|||||||||:||||::::|||
Db  251 GAACCCAGATACACACAGATGCACATGAGGAGGCGGTTTGAAGCATGGGTGCCTGCAGGAAAG 310

Qy   65 SerProThrSerGluLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeu 84
        |||||||||||||||||
Db  311 AGTCCACCATGGAGCTGCTGTTCGACTGGGGAACCACCTCAGACTGCACTGTTGGAGATCTG 370

Qy   85 ValAspLeuLeuIleGlnAsnGluPheAlaProAlaSerLeuLeuLeuProAsp---- 103
        :::|:::||||||||::::
Db  371 GTGGAGATCCTGATCCGCCATCAGCGTGTTCGCGCGCCGTCACAGTCCTGCCGATCAC 430

Qy  104 --------------------------------AlaValProLysThrAla 109
                                        ||| :::|||
Db  431 AGGGTCTGCCAGCTCCACGGTCACCAGGCTCACAGGTGGAGGATGAAAACCGGTCACAGGAGCAG 490

Qy  110 AsnThrLeuProSerLysGluAlaIleThrValGlnGlnLeuGlnMetProPheCysAsp 129
        :::|||
Db  491 GTGTGTCTG--------CAGGCATGCGAAAATCACCCAAAACTGTTAGAGGACAACAGTAAT 541

Qy  130 LysAspArgThrLeuMetThrProValGlnAsn--------------------LeuGlu 142
        ||||||:::||:::
Db  542 AAACCCCAAAAAATCTCTAAACCTGTGGAGGATGACAGTAATAAAACCGGTCACAGGAGCAG 601

Qy  143 GlnSerTyrMetProProAspSerSerProGluAsnLysSerLeuGluValSerAsp 162
        ||||||:|:|||
Db  602 GAGCTGTTTGTGGAGCCGGACACAGCGTTCTGGAGACACAGGAGGCAGCTGGAACTCCAGC 661

Qy  163 ThrArgPheHisSerPheTyrGluLeuLysAsnValThrAsnAsnPheAspGlu 182
        :::|||||:|:||
Db  662 CAAGGCTTCCACACCTTCAGTCCTCCATGAGCTGACGCGGGCCGATCAGCCCAGGAGCAG 721

Qy  183 ArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyValValTyrLys 202
        |||||||:::||||
Db  722 AGGGCACTTCGGATGGGGTGGGGTGCTGCCGCTCGGAGGGTTCGGAGTGGTGGTCAGA 781

Qy  203 GlyTyrValAsnAsnThrThrValAlaValAlaValLysLysLeuAlaAlaMetValAspIleThr 222
        ||||||||||||||||
Db  722 :::  :::
```

```
BASE COUNT      168 a     94 c    111 g    163 t      3 others

FEATURES             Location/Qualifiers
     source          1..539
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"

SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Obata,Y.
  TITLE     Breast, gastric and prostate cancer associated antigens and uses
            therefor
  JOURNAL   Patent: WO 0073801-A 197 07-DEC-2000;
            LUDWIG INSTITUTE FOR CANCER RESEARCH (US)
KEYWORDS    .
VERSION     AX053431.1  GI:12227750
ACCESSION   AX053431
DEFINITION  Sequence 197 from Patent WO0073801.
LOCUS       AX053431              539 bp    DNA       linear   PAT 13-JAN-2001

RESULT 15
```

```
Qy  1391 CAAGAGACCCGCCATCGCTCAGGTTCTGTCGTCGTGGAGGATCTG 1436
         :::|||||||||||||||||||||||:::||||:::::|||||||
Db  1391 nLysArgProAspIleLysLysValGlnGlnLeuLeuGlnGluMet 457
```

```
Qy   442 nLysArgProAspIleLysLysValGlnGlnLeuLeuGlnGluMet 457
         |||||||||||||||:::||||
Db   442 pSerThrSerValGluAlaMetTyrSerValAlaSerGlnCys 442
Qy  1331 AACAGAAGAGCTGCAGATCATGTATGAAGCGGGCGTTCAGTGTCTTCGTGCCAGAAGAGAA 1390
         :::       |||||||||||||||||||||||||||||||||||||||||||
Db  1331 pSerThrSerValGluAlaMetTyrSerValAlaSerGlnCys 442
Qy   422 pSerThrSerValGluAlaMetTyrSerValAlaSerGlnCys 442
         |||||||
Db   422 uIleGluGlyAspGluGlyLysThrIleAspLysTyrIleAspGlySly 422
Qy  1271 TCTCGATGATGAGGATCTCTCTCGTGCTGGACTTCACTGACCGGCCCATGCAGACTGGAG 1330
         :::|||||||||||||||||||||||||||||||||||||||||
Db  1271 uIleGluGlyAspGluGlyLysThrIleAspLysTyrIleAspGlySly 422
Qy   402 uIleGluGlyAspGluGlyLysThrIleAspLysTyrIleAspGlySly 422
         ::||||||||||||||||||||:::||||
Db   402 rGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeu 402
Qy  1211 TGGGCCTTCCTTCCAGTGGACGAGAGTCGGAGGAACCCTGCTTGCTGGATGAAGGATGA 1270
         :::|||||||||||||||||||||||||||||||:::|||:::||
Db  1211 rGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeu 402
Qy  1178 rGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeu 402
         ------------------TCTGCGCGCCTGCAGGTGCTGCAGGAGGTTTTGTC 1210
Db  1178 -------------------------------------TCTGCGCGCCTGCAGGTGCTGCAGGAGGTTTTGTC 1210
Qy  1177 ---------------------------------------------------- 1177
Db  1177 ---------------------------------------------------- 1177
Qy   363 GluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIleIleTh 382
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 GluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeu 382
Qy   283 LeuSerTrpHisMetArgCysLysIleAlaGlnAlaAsnGlyIleAsnPheLeu 302
         ||||||||||||||||||||||||||||||||||||||||||:::||||||
Db   283 LeuSerTrpHisMetArgCysLysIleAlaGlnAlaAsnGlyIleAsnPheLeu 302
Qy  1022 CTCACCTGGAGGACAGATGCTGGATCACTGTTGGAGTCGGCCGGGGCCTGAGCTACCTG 1081
         :::||||||||||||||||||||||||||||||:::||||||||||||||||
Db  1022 LeuSerTrpHisMetArgCysLysIleAlaGlnAlaAsnGlyIleAsnPheLeu 302
Qy   303 HisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAspGluAla 322
         |||||||||||||||||||||||||||||||||||||:::||
Db   303 HisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAspGluAla 322
Qy  1082 CACACACACGCACACATCCACAGAGACGTCAAGAGCGCCAACACTCTTACTGCAGAGGCT 1141
         :::
Db  1082 GlyIleAsnPheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIle 317
Qy   298 GlyIleAsnPheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIle 317
         ||||||||||||||||||||||||||||||||||||||||||||
Db   298 GGCATCAATTTTCTACAGAAAATCATCATATTTCACATGAGATGCATAGATGCAAGATG 317
Qy   278 AspGlyThrProProLeuSerPheHisMetArgCysLysIleAlaGlnAlaAsn 297
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   278 AspGlyThrProProLeuSerPheHisMetArgCysLysIleAlaGlnAlaAsn 297
Qy    61 GATGGGTACTCCACCACTTTCTTTGGACATGAGATGCAAGATGCTGCAGGTGGC 120
         ||||||||||||||||||||||||||||||||||||||||||||
Db    61 AspGlyThrProProLeuSerPheHisMetArgCysLysIleAlaGlnAlaAsn 120
Qy   323 PheThrAlaLysIleSerAspPheGlyLeuAlaArgAspAlaSerGlyHisPheAlaGln 342
         ||||||||||||||||||||||||||||||||
Db   323 PheThrAlaLysIleSerAspPheGlyLeuAlaArgAspAlaSerGlyHisPheAlaGln 342
Qy  1142 TTTGTGCGGAAGACTCAGACTTCGGGCTTGACTCGG 1177
         ||||
Db  1142 LeuLeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAspAlaSerGlu 337
Qy   343 ValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeuArgGly 362
         |||
Db   343 ValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeuArgGly 362
Qy   263 ValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThrProPro 282
         |||||||||||:::|||||||:::         :::
Db   263 ValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThrProPro 282
Qy   962 GAGCTGATGGTGAACGGCTCTCTGCTGGACGAACTGGTCTGCCTTGGGCGTGAACACACCTGCG 1021
         :::
Db   962 LeuLeuAspArgLeuSerCysLeuAspGlyThrProProLeuSerPheHisMetArg 1021
Qy   243 GluAsnLeuValGluLeuGlyPheSerSerAspGlyAspAspIleCysLeuValTyr 262
         ::::::||||||||||||||||||:::||:::|||
Db   243 GluAsnLeuValGluLeuGlyPheSerSerAspGlyAspAspIleCysLeuValTyr 262
Qy   902 GAGAATGTGTTGCGTGTGTTGGGCGTGCTGCTCCGGCGCCCGCCTGTGTGTGTGTTT 961
         :::::::
Db   902 GAGAATGTGTTGCGTGTGTTGGGCGTGCTGCTCCGGCGCCCGCCTGTGTGTGTGTTT 961
Qy   223 ThrGluGluLeuLysGlnLnPheAspGlnGluIleLysValMetAlaLysCysGlnHis 242
         ||:::|||||||||:::||||||
Db   223 ThrGluGluLeuLysGlnLnPheAspGlnGluIleLysValMetAlaLysCysGlnHis 242
Qy   842 TATGAAGACCTCAGAAAGCAGTTTAACCAGACACTCAGAAGTCTGAGTCAT 901
         |||
Db   842 ThrGluGluLeuLysGlnLnPheAspGlnGluIleLysValMetAlaLysCysGlnHis 841
Qy   782 GGACCCATGGGAGGGTAAACATGTGGCGCGGTCAAGAAACTCAATCCTTGGACGGCAGTTCA 841
Db   782 ---
```

```
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:           883.00       Matches:          176
Percent Similarity:  97.24%   Conservative:       0
Best Local Similarity:  97.24%  Mismatches:       3
Query Match:     37.01%        Indels:            2
DB:              6             Gaps:              0
                                Length:         539
                 6.4e-68
```

```
US-10-001-254-16 (1-460) x AX053431 (1-539)
```

```
Qy   258 LeuCysGluValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeu 277
         :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTCTGCTTAGTATATGTTTACATGCCTAAATGGTTCATTGCTTAGACAGAGACTCTCTTGCTTG 60
Qy     1 AspGlyThrProProLeuSerPheHisMetArgCysLysIleAlaGlnAlaAsn 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GATGGGTACTCCACCACTTTCTTTGGACATGAGATGCAAGATGCTGCAGGTGGC 120
Qy   298 GlyIleAsnPheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIle 317
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGCATCAATTTTCTACAGAAAATCATCATATTTCACATGAGATGCATAGATGCAAATATC 180
Qy   318 LeuLeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAspAlaSerGlu 337
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TTACTGGATGGAAGCTTTTACTGCTAAAAATATCTGACTTTGGCCTTGCACGGGCTTCTGAG 240
Qy   338 LysPheAlaGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaPro 357
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AAGTTTGCCCAGACAGTCATGACAAGAATTGTGGGAACAACAGCTTATATGGCACCA 300
Qy   358 GluAlaLeuArgGlyValGluIleThrProLysSerAspIleTyrSerPheGlyValValLeu 377
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GAAGCTTTGCGTGGAGAAATAACACCCAAAATCTGATATTTACAGCTTTGGTGTGGTTTTA 360
Qy   378 LeuGluIleIleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeu 397
         |||||||||||||||||||||||||||||||||||||||||||:::||||:::||
Db   361 CTAGAAATAATAAACTGGACTTCCAGCTGTGGATGAACACCGGGAACACCCGGGAACACC 420
Qy   398 AspIle---LysGluGlyAspGluGlyLysThrIleAspLysTyrIleAspGlySly 417
         |||||||                     :::|||||||||||
Db   421 GATATTAAAGAAGGAGAAATTGNAGATGAAGAAGACAATTGAAGAGATTATTATGATAAAAA 480
Qy   417 sMetAsnAspAlaAspSerThrSerValGluAlaMetTyr-SerValAlaSerGlnCys 436
         |||||||||||||||||||||||||||||||||||||||          :::
Db   481 GATGAATGAATGCTGATTCCACTTCAGTTGAANCTANGTACCTCTGTTGNTAGTCAATGT 539
```

```
Search completed: January 18, 2004, 04:26:40
Job time : 7811.21 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 18, 2004, 01:16:08 ; Search time 120.032 Seconds
(without alignments)
1691.512 Million cell updates/sec

Title: US-10-001-254-16
Perfect score: 2386
Sequence: 1 MNKPITPSTYVRCLNVGLIR..........KNKRPDIKKVQQLLQEMTAS 460

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000.

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10001254//runat 16012004_152425_19740/app_query.fasta_1.1109
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10001254 @CGN 1_1_95 @runat 16012004 152425 19740 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1169	49.0	833	4	US-09-166-350-10	Sequence 10, Appli
2	515.5	21.6	3590	1	US-08-587-889-1	Sequence 1, Appli
3	515.5	21.6	3590	5	US-09-016-434-1093	Sequence 1093, Ap
4	515.5	21.6	3590	5	PCT-US96-09193-1	Sequence 1, Appli
5	476.5	20.0	1488	4	US-09-579-182-4	Sequence 4, Appli
6	452.5	19.0	2336	4	US-09-228-986-10	Sequence 10, Appl
7	451	18.9	3239	4	US-09-228-986-9	Sequence 9, Appli
8	448.5	18.8	2432	4	US-09-228-986-7	Sequence 7, Appli
9	438	18.4	2288	3	US-09-135-232-1	Sequence 1, Appli
10	438	18.4	2288	4	US-09-863-549-1	Sequence 1, Appli
11	432	18.1	2868	4	US-09-228-986-4	Sequence 4, Appli
12	430.5	18.0	966	1	US-08-447-185-2	Sequence 2, Appli
13	430.5	18.0	2443	1	US-08-447-185-3	Sequence 3, Appli
14	427.5	17.9	2749	2	US-07-717-331F-4	Sequence 4, Appli
15	424	17.8	3459	3	US-08-980-060-3	Sequence 3, Appli
16	424	17.8	3459	5	US-09-307-185-3	Sequence 3, Appli
17	423	17.7	1806	4	US-08-980-060-1	Sequence 1, Appli
18	423	17.7	1806	3	US-09-307-185-1	Sequence 1, Appli
19	410.5	17.2	2943	4	US-09-503-922-2	Sequence 2, Appli
20	403	16.9	283	1	US-09-016-434-998	Sequence 998, App
21	402.5	16.9	2749	1	US-08-265-628-1	Sequence 1, Appli
22	398.5	16.7	2571	1	US-07-717-331F-1	Sequence 1, Appli
23	398.5	16.7	2833	1	US-09-579-182-3	Sequence 3, Appli
24	391.5	16.4	3097	1	US-07-717-331F-9	Sequence 9, Appli
25	390.5	16.4	1224	2	US-08-587-680A-24	Sequence 24, Appli
26	386	16.2	1554	2	US-08-881-706-1	Sequence 1, Appli
27	386	16.2	4104	3	US-09-228-986-1	Sequence 1, Appli
28	382.5	16.0	2389	4	US-09-228-986-13	Sequence 13, Appli
29	379	15.9	5733	2	US-08-473-553A-1	Sequence 1, Appli
30	377	15.8	2513	2	US-09-228-986-13	Sequence 13, Appli
31	358.5	15.0	2686	4	US-09-228-986-3	Sequence 3, Appli
32	341.5	14.3	2510	2	US-08-888-982A-42	Sequence 42, Appli
33	341.5	14.3	2510	4	US-09-462-261-42	Sequence 42, Appli
34	341.5	14.3	2510	4	US-09-506-073-89	Sequence 89, Appli
35	337.5	14.1	2638	4	US-09-228-986-8	Sequence 8, Appli
36	321	13.5	1947	1	US-08-185-282-11	Sequence 11, Appli
37	321	13.5	2977	1	US-08-276-151-1	Sequence 1, Appli
38	321	13.5	2977	2	US-08-306-691B-17	Sequence 17, Appli
39	321	13.5	2977	3	US-08-328-233A-4	Sequence 4, Appli
40	321	13.5	2977	2	US-08-756-806A-64	Sequence 64, Appli
41	321	13.5	2977	3	US-09-143-214-64	Sequence 64, Appli
42	321	13.5	2977	4	US-09-205-668-12	Sequence 12, Appli
43	321	13.5	2977	4	US-09-209-668-25	Sequence 25, Appli
44	321	13.5	2977	4	US-09-555-279-1	Sequence 1, Appli
45	321	13.5	2977	5	PCT-US95-13661-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-166-350-10

; Sequence 10, Application US/09166350A
; Patent No. 6440663
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Chen, Yao
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alex
; TITLE OF INVENTION: Renal Cancer Associated Antigens and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: L0461/7051
; CURRENT APPLICATION NUMBER: US/09/166,350A
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: US 09/166,350
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 10
; LENGTH: 833
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-166-350-10

Alignment Scores:
Pred. No.: 1.31e-127
Score: 1169.00
Percent Similarity: 93.21%
Best Local Similarity: 92.83%
Query Match: 48.99%

DB:

	Length:	833
	Matches:	246
	Conservative:	1
	Mismatches:	8
	Indels:	10
	Gaps:	3

US-10-001-254-16 (1-460) x US-09-166-350-10 (1-833)

GENERAL INFORMATION:
APPLICANT: CAO, Zhaodan
APPLICANT: CROSTON, Glenn E.
APPLICANT: GOEDDEL, David V.
TITLE OF INVENTION: INTERLEUKIN-1 RECEPTOR-ASSOCIATED
TITLE OF INVENTION: PROTEIN KINASE AND BINDING ASSAY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/587,889
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-60916
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3590 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-587-889-1

RESULT 2
US-08-587-889-1
Sequence 1, Application US/08587889
Patent No. 5654397

US-10-001-254-16 (1-460) x US-08-587-889-1 (1-3590)

Query Match:
Best Local Similarity: 21.61%

Alignment Scores:
Pred. No.:   6.85e-50
Score: 515.50
Percent Similarity: 46.14%
Best Local Similarity: 29.38%
Query Match: 21.61%

Length: 3590
Matches: 156
Conservative: 89
Mismatches: 192
Indels: 94
Gaps: 19

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
Db   4 ProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSer 23

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys 40
Db   24 AspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaMetAlaVal 43

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
Db   44 GlyAspAspArg---TyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGl 63

Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
Db   63 TyrLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAs 83

Qy   81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPheAlaProAlaSerLeuLeu 100
Db   93 pLeuValAspLeuLeu---IleGlnAsnGluPh 93

Qy 101 LeuProAspAlaValProLysSerThrLysThrLeuProSerLeuGluValValVal 120
Db 115 ProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSer 174

Qy 121 GlnGlnLysGlnMetProPheCysAspLysLysAspArgThrLeuMetThrProValGlnAsn 140
Db 175 AGT-----GAtGGAGCGCCCTGGAGCCCGCCACTG---GTCCCAGTTCGCGCCCT--- 222

Qy 141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerSerLeuGluVal 160
Db 223 ------GATCGTGCCGCGGACCAGACCGAGCTGCGG------CTGTGCGAGCGCTC 264

Qy 161 SerAspThrArgPheHisSerPheSerPheTyrGluLysAsnValThrAsnAsnPhe 180
Db 265 CGGGCAGCGCGCACGGCGCTGGATCAACCGACCAACCGCCGTGTGGCCGA 324

Qy 181 AspGluArgProIleSerGluValGlyGlyAsnLysSerMetGlyValGluGlyGlyGlyValVal 200
Db 325 CCTCGTCCACATCCTCACGCACCTGCAGCTGCTCCGTGCCGCGGGCGACATCATCACAGCCTG 384

Qy 201 Tyr-LysGlyTyrValAspAsnThrThrArgValAlaVal---LysLysLeuAlaLaMetValA 220
Db 385 GCACCCTTCCCCGGCGCCCCGGCCTTCCGTCCCAGGGACCACTGCCCCGGCCAGCAGCATCC 444

Qy 220 spIleThrThrGlu-GluLeuLys-GlnInPheAsp-GlnGlu-IleLysValMetAla 238
Db 445 TGCACCCGCCCGAGGCGCCTGGAGCGCCCCGGAGTTGCCA------TCCTCAGCGTC 498

Qy 239 LysCysGlnHis------GluAsnLeuVal---GluLeuLeuGlyPheSerSer---Asp 254
Db 499 CACCTTCCTCTCCCCAGCTTTTCCAGGACGCCCCATTCAGGGCGGTGAGCTCGGCCT 558

Qy 255 Gly 255
Db 559 GGTTCCAAGCCCCCTGCTTCCCCGTGGGCCTCCACCGCCATCTCCAGCGCCCTTCTACCAA 618

Qy 255 Gly 255
Db 830 GGG 832

RESULT 3

```
Qy    157 r-------------------------LeuGluValSerAspThrArgPheHisSerPh 168
                                   :: :::
Db    619 GCCAGGCCCAGAGAGCTCAGTGTCCTCCTGCAGGGAGCCGCCCCTTCCGTTTGCTG 678

Qy    168 eSerPheTyrGluLeuTyrAsnValThrAsnAspPheAspGluArgProIleSerValGl 188
              ||:::                   ||::
Db    679 GCCCCTTGTGAGATTTCCGGGGCACCCAACTTTCTCGGAGGAG------------ 724

Qy    188 yGlyAsnLysMetGlyIleGlyPheValValIleTyrLysGlyTyrValAsnAsnTh 208
                                        ::
Db    725 ------CTCAAGATCGGGAGGGTGAGGGTGGCTTTGGGTGCGTGTCGATGAGGAACAC 780

Qy    208 rThrValAlaValLysLysGlyLeuAlaAlaMetValAspIleThrThrGluGluLysGl 228
              ||||||||:::::||
Db    781 GGTGTATGCTGTGAAGAAGGGCTGAAGGAGGAGAAACGCTGAGGGGACTGCAGTGAAGCA 840

Qy    228 nGlnPheAspGlnGluIleLysValMetAlaLysCysGlnHisGluAsnLeuValGlu 248
              |||||||::                     ::||||:::||:::
Db    841 GAGCTTCCTGACCGAGGGTGGAGCAGCTGTCCAGGTTTGTCACCCAAACATTGTGGACTT 900

Qy    248 uLeuGlyPheSerSerAspGlyAspLeuCysLeuValTyrValTyrMetProAsnGl 268
                                           :::
Db    901 TGCTGGCTTACTGTCGTCCAGAACCGGCTTCTACTCGTTCCTGGTACGGCTTCCTGCCAA 960

Qy    268 yGerLeuLeuAspAspArgLeuSerCys----LeuAspGlyThrProProLeuSerTrpHisMe 287
              ||:::||||::                 |||||::
Db    961 CTCCCTGGAGGACCGTCTCCACTGCCCAGACCCGGCAATTCAGTTTCTACATCAGACAGCCC 1020

Qy    287 tArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleLeuAsnPheLeuHisSerGluAsnHisHi 307
              ||||||::||||                  :||||||||:::::
Db   1021 GGCACTGACATCCTTCTGGGTACAGCCCGGCAAGTCCTTCCTGCCAACGG 1080

Qy    307 s-------IleHisArgAspPheLysSerAlaAsnIleLeuLeuAspGluAlaPheThrVal 325
                       ||||||||||||
Db   1081 CAGCCTCATCCAATGGAGCAGCATCAAGAGTTCCAAGGTCCTTCTCTGGATGAGAGGGCTGACACC 1140

Qy    325 aLysIleSerAspPheGlyLeuAlaArgAspAlaSerGlyLysSerPheAla----------- 340
              ||||||||||::||
Db   1141 CAAGCCTGGGAGACTTTTGGCCTGGCCCGGTTCAGC---CGCTTTGCCGGGGTTCAGCCCCAG 1197

Qy    341 -GlnThrValMetThrSerArg------------IleValGlyThrThrAlaTyrMetAlaPr 357
                     :::                    ||||||||
Db   1198 CCAGAGCAGCATGGTGGCCCGGCACACAGACAGTTGCGGGGGCCTACCTGCCGA 1257

Qy    357 oGluAlaLeuArg----GlyGluIleThrProLysSerAspIleThrSerProPheGlyValVa 376
              ||:||||                   ||:::||||||||||||
Db   1258 GGAGTACATCAAGACAGCGGAAAGCCTGGCTGTGGACACCAACCTTCAGCTTTGGGGTTGGT 1317

Qy    376 lLeuLeuGluIleIleThrGlyLeuProAlaValAspGluHis---ArgGluProGlnLe 395
              |||||||||||                       ::||
Db   1318 AGTGCTAGAGACCTTGGCTGGTCAGAGGGCTGTGAAGACGCACGGGTGCCCAGTTGCCA 1377

Qy    395 uLeuLeuLysAspIleLysGluIleLeuGluAspGlu------------Gl 407
              ||||||||:::||||::                          ::
Db   1378 TCTGAAAGACCACGTTGTGAAGAGAAGAGGGCTGAGAGTGGCTTTGAGTAAGCACCCA 1437

Qy    407 uLysThrIleGlu--------- 411
              :|||
Db   1438 GAGCACACTGCAAGCAGGTCTGGCTGCGACATGCCCATGCCATGCAGAT 1497

Qy    412 -------AspThrIleAspLysLysMetAsnAspSerThrSerValGluAlaMe 429
                     :::|||||
Db   1498 CTACAAGAGCACCTGGACACCCAGGCCCTGCCCACTCGAGCTGTGGGCGCCGCCT-- 1557

Qy    429 tTyrSerValAlaSerGlnCysTrpLeuHisGluLysLysAsnLysArgProAspIleLysLy 449
              |||||                      :|||::
Db   1558 GGGGCACCTGGCCGCTGCTGCTGCTGCTGCTGCTCCGCCGCCGCCTCCTTATGACCCA 1617

Qy    449 sValGlnGlnLeuLeuGluGlnGluMetThrAla 459
              :|||     :||:::||:
Db   1618 GGTGTACCGAGAGAGGCTAGAGAAGCTGCAGGCA 1648
```

Right column:

```
US-09-016-434-1093
; Sequence 1093, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1093:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3590 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 91220312
US-09-016-434-1093

Alignment Scores:
Pred. No.:
Score:                      6.85e-50
Percent Similarity:         515.50      Length:    3590
Best Local Similarity:      46.14%      Matches:   156
Query Match:                29.38%      Conservative: 89
DB:                         21.61%      Mismatches: 192
                            4           Indels:    94
                                        Gaps:      19

US-10-001-254-16 (1-460) x US-09-016-434-1093 (1-3590)
```

```
Qy      4 ProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgLysLeuSer 23
             ||||          ::                 :::
Db    115 CCCCGGCGCCCACCACTTCTTGTACGAGGTGCGCGCCTTGGGTCATGTGCCGCTTCTACAA 174

Qy     24 AspPheIleAspProGlnGlyIleTrpLysLysLeuAlaValAlaIleLysLysProSer 43
             |||           :::::|||    ::       ||     ||
Db    175 AGT------GATGGACGCCCCTGGAGCCCCGCGACTG--GTTGCCAGTTGCCGCCCCCT-- 222

Qy     44 GlyAspAspArg-TyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGl 63
             |||||||         |||||           :::
Db    223 ------GATCGTCGGCGACCAGACCGAGCTGCGCG-- 264

Qy     63 yLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAs 83
             :|||          ::        |||::               :::
Db    265 CGGGCAGCGCACGGCGCCACGGTCCTGTGGCCCTGGGATCAACCCGGACACCCCGGTGTGGCCGA 324
```

RESULT 4
PCT-US96-09193-1
; Sequence 1, Application PCT/US9609193
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Tularik, Inc.
; TITLE OF INVENTION: INTERLEUKIN-1 RECEPTOR-ASSOCIATED
; TITLE OF INVENTION: PROTEIN KINASE AND BINDING ASSAY
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09193
; FILING DATE: JUNE 5 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Serial No. 08/587,889
; FILING DATE: JAN 16 1996
; CLASSIFICATION:
; APPLICATION NUMBER: U.S. Serial No. 08/494,006
; FILING DATE: JUNE 23 1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: David J. Brezner
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: FP-62191-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3590 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US96-09193-1

Alignment Scores:
Pred. No.:   6.85e-50
Score: 515.50   Length: 3590
Percent Similarity: 46.14%   Matches: 156
Best Local Similarity: 29.38%   Conservative: 89
Query Match: 21.61%   Mismatches: 192
                  Indels: 94

DB: US-10-001-254-16 (1-460) x PCT-US96-09193-1 (1-3590)

Gaps: 5   19

```
Qy      4  ProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSer      23
           |||::::                              ::
Db    115  CCCGGCCCCAGCACTTCTTGTACGAGGTGCCGGCCCTGGGTCATGTGCCGGTTCTACAA     174

Qy     24  AspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSer      43
           :::                              ::::
Db    175  AGT-------GATGGACCCCTGAGCCCCGCCACTG--GTGCCAGTTCGGCCCCT---     222

Qy     44  GlyAspAspArg-TyrAsnGlnPheHisIleArgArgPheGluAlaLeuGlnThrGl      63
                                                                     ::
Db    223  ------------GATCGTGCGGCCGACCAGACAGGCAGGC------CTGTGCGAGCCGCTC     264
```

RESULT 5
US-09-579-182-4
; Sequence 4, Application US/09579182
; Patent No. 6500628
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE AND
; TITLE OF INVENTION: PHOSPHATASE HOMOLOGUES AND USES THEREFOR
; FILE REFERENCE: MNI-161
; CURRENT APPLICATION NUMBER: US/09/579,182
; CURRENT FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1488
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-579-182-4

Alignment Scores:
Pred. No.:
Score: 6.19e-46
                           476.50
Percent Similarity: 54.01%
Best Local Similarity: 38.58%
Query Match: 19.97%
DB: 4

Length: 1488
Matches: 130
Conservative: 52
Mismatches: 104
Indels: 51
Gaps: 14

US-10-001-254-16 (1-460) x US-09-579-182-4 (1-1488)

RESULT 6
US-09-228-986-10
Sequence 10, Application US/09228986
Patent No. 6359198

US-10-001-254-16 (1-460) x US-09-228-986-10 (1-2336)

Alignment Scores:
Pred. No.:
Score: 452.50   Length: 2336
Percent Similarity: 51.93%   Matches: 124
Percent Identity: 36.80%   Conservative: 51
Best Local Similarity: 18.96%   Mismatches: 115
Query Match: 4   Indels: 47
DB: Gaps: 13

(Sequence alignment data — paired Qy / Db rows with position numbers, match bars (|||:::), amino-acid translations and nucleotide sequences; e.g.:)

Qy   112 ACTGCTCCAAAAGAAGGGCCTACTGCGCATATTCGTCGCACAAACCTTTACTTTCGACGAG 171
Db   173 ...

Db   173 LeuLysAsnValThrAsnAsnPheAspGluArgProIleSerValGlyGlyAsnLysMet 192
Db   172 TTAGCTGCCGGCCACTAAAAACTTT----CGACCGGAAATGTCTT-- 213

Qy   211 AlaValLysLeuValAlaAlaMetValAspIlePheThrThrGlyGluLeuLysGlnGlnPhe 230
...

Qy   434 SerGlnCysLeuHisGluLysLysSerAsnLysArgProAspIleLysGlyLysVal 450
Db   955 GCAATGTGTTTACACGGAACAAGCACGACAAGACCACTGATTGGGCGACGTG 1005

Db   383 GlyLeuProAlaValAsp-----GluHisArgGluProGlnLeuLeuLeuAspIleLys 400
Db 1591 AAGGCAACTCCTAGCTCCGGACACGTGTTCAGCTTCGGTTCGGCTTGC 1650

RESULT 7

US-09-228-986-9

; Sequence 9, Application US/09228986

; Patent No. 6359198

; GENERAL INFORMATION:

; APPLICANT: Strabala, Timothy

; APPLICANT: Nieuwenhuizen, Niels

; TITLE OF INVENTION: Compositions Isolated from Plant Cells

; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling

; FILE REFERENCE: 11000/1020

; CURRENT APPLICATION NUMBER: US/09/228,986

; CURRENT FILING DATE: 1999-01-12

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 9

; LENGTH: 3239

; TYPE: DNA

; ORGANISM: Eucalyptus grandis

US-09-228-986-9

US-10-001-254-16 (1-460) x US-09-228-986-9 (1-3239)

Alignment Scores:

Pred. No.:
Score: 451.00
Percent Similarity: 47.17%
Best Local Similarity: 28.95%
Query Match: 18.90%

Length: 3239
Matches: 143
Conservative: 90
Mismatches: 159
Indels: 102
Gaps: 23

Qy 15 AsnValGlyLeuIleArgGlyLeuSerAspPheIleAspProGlnLeuGlyTrpLysLys 34
        |||||| |||| | || ||| ||| |
Db 1555 AAtGTGACCCTTGTTTCACATGCTCGAA------------------ATGGACTTGTGGACGAAG 1599

Qy 35 Leu---------------AlaValAlaIleLeuLysLysProSerGlyAsp---AspArgTyr 48
        ::: ||| || || ||| ::: |||||
Db 1600 TTGAATCTCACTCCCCGGTTCGGTTTCTTCAGAACCCTTCTCAATCTTGACGACTAC 1659

Qy 49 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLys----------- 64
        ::||| ||| ||| || ::
Db 1660 CTTCAAGTCGCAGCTTTCACTCTTCCCC------------CCGAGCGGGAAATATTTCAGTCGG 1710

Qy 65 SerProThrSerGlyLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeu 84
        ||||| | ||| | |||
Db 1711 TCGGATATTCAGAGACCATCGGTTTCGAC------TATGTCGTCCGGCAGACG------GACAGAACGAC 1737

Qy 85 ValAspLeuLeuGlnAsnGluPhePheAlaProAlaSer------------ 98
        ||| ::: || :::
Db 1738 TTGACAACCAACTTTCAAGCCTCCAAACCATTCGGCCCCTATTAC 1785

Qy 99 ---LeuLeuProAspAlaValProLysThrLeuProSerLysGluAla 117
        ::: ||||| ||| |||||| ::
Db 1786 TTCATGCGCTCCCCCTATGCTTTCCAGACAAGCCCCATAGCCAAAGGTGTG 1845

Qy 118 Ile   118
        |||
Db 1846 ATAGTTGGGATGCGTATTGGCGGCACGGTTCTCGGTTCTTGGCCTTGTGTATTAGGGTTA 1905

RESULT 8

Qy 119 ---ThrValGlnGlnLysGlnMetProPheCysAspGlyLysAspArgThrLeuMetThrPro 137
        :::||||| ||| ||||||| ::|||
Db 1906 TATGCTATTCGACAAAAGAAACGG---------GCCGAGAAGCTCTCGAGTTGACAGACCC 1959

Qy 138 ValGlnAsnLeuGluGlnSerTyrMetProProAspSerSerProGluAsnLysSer 157
        |||||| || ::   :::
Db 1960 TTC-----------GCAtCCTCGGGCACCC------AGTGGGAAAGATAGCCGAGGAGCG 2001

Qy 158 LeuGluValSerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThr 177
        ::|||| ||
Db 2002 CCACAACTGAAAGGACGACGATGG------TTTCCTATGATGAACTTAAGAGGTGCACC 2055

Qy 178 AsnAsnPheAspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPhe 197
        || :::   :: ::
Db 2056 AATAATTTTCTCCGAT------AGCAATGAAATTAGGCTTCGGAGGATAC 2097

Qy 198 GlyValValTyrTyrLysGlyTyrTyrVal---AsnAsnThrThrValAlaValAlaValLeu--- 215
        |||||| :::   :::
Db 2098 GGAAAGGTGTACAGCGGGAGTTCTTCCCTGATGGTCCATATATTACCAATCAAAGAGCTCAG 2157

Qy 216 ---AlaAlaMetValAspIleThrThrGluGluLeuLysGlnPheGlnGlnGluIle 234
        ::: ||
Db 2158 CAGGGGTCGAtCGCAGGGTGCACC------GAGTTCAAGACAGAAATC 2199

Qy 235 LysValMetAlaLysCysGlnHisGluAlaAsnLeuValLeuGluLeuGlyPheSerSerAsp 254
        |||||| || ::: :::|| :
Db 2200 GAGCTGCTTCGGGGGTTCATCACAGAATCTTGTTGGCCCTCATAGGATTCTGTTTCGAG 2259

Qy 255 GlyAspAspLeuCysLeuValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeu 274
        :::
Db 2260 CAAGGAGACAGATGTTGGTCTTCATGAATATATGCCTAACGGATCGCAGGGATTAGCTTG 2319

Qy 275 SerCysLeuAspGlyThrProLeuSerTrpHisMetArgCysLysIleAlaGlnGly 294
        :::
Db 2320 ACAGGAAAATCAGGCATT---TATCTTCAATTGGAAGAGAGGACTTCGTATAGCTCTAGGT 2376

Qy 295 AlaAlaAsnGlyIleAsnPheLeuHisGluAsnHisHis------------IleHisArgAsp 311
        ||| ::: ||||||
Db 2377 TCGGCTAGAGGACTAGCTTATCTGAGTCAACTCCAATATCCACAGAGAT 2436

Qy 312 IleLysSerAlaAsnIleLeuLeuAspGluAlaLeuAlaProPheThrAlaLysIleSerAspPheGly 331
        ::: |||||| :||||
Db 2437 GTCAAGTCCACCAATATCTTGTTGGACGAACATCTGACGGCCAAAAGTCGCCGGATTTCGGT 2496

Qy 332 LeuAlaArgAlaSerGluLysPheAlaGlnThrValMetThrSerArgIleValGlyThr 351
        |||
Db 2497 TTGTCCAAACAGGTATCCGGCAGAGGGGCCACGTTTCGACGCGAAGTGAAAAGTCACG 2556

Qy 352 ThrAlaTyrMetAlaProGlu---AlaLeuArgGlyValGluIleThrProLysSerAspIle 370
        ::: |||| ||| ||
Db 2557 CTCGGGCTATTTGGATCCCGGAATAACTACAGTCAACAGCTGACAGAAAAGACCGGATGTG 2616

Qy 371 TyrSerPheGlyValValLeuLeuGluIleLeuIleThrGlyLeuProAlaValAspGluHis 390
        |||||| |||||| :||||||
Db 2617 TACAGCTTCGGGGTGGTCACGCTTCATCATCGGGATCGCAAGCACCGATGAGAGGC 2676

Qy 391 ArgGluProGlnLeuLeuAspIleLysSerGlyLeuAspGlyLysThrIle 410
        ||| ::: |||||||
Db 2677 AAG---------TATGTCGTCCGGCAGAGGGGCCATG------GACAGAACGAC 2721

Qy 411 GluAspTyrIleAsp---LysLysMetAsnAspAlaAspSerThrSerVal--------- 426
        :::||||||| ||
Db 2722 CAGGACTACTACCGGCGTGAGGAAATGATGGACCGTCCATGAGCCGGCTACTTC 2781

Qy 427 ------GluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLysLysAsnLys 443
        ::: |||
Db 2782 GTCGGGGTTCAGCAGGTTCTTGGATTTGGCGATCGGAGGATGTCGGAGAGTCGGCGGAC 2841

Qy 444 ArgProAspIleLysLysValGlnGlnLeuGluGluGlyMet 457
        |||
Db 2842 CGCCCCACAATGAGCAGAGGGTTGCGTGAAGGCGGATCGAGACCATG 2883

US-09-228-986-7
; Sequence 7, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2432
; TYPE: DNA
; ORGANISM: Pinus radiata

US-10-001-254-16 (1-460) x US-09-228-986-7 (1-2432)

Alignment Scores:
Pred. No.:
Score: 2.6e-42
Percent Similarity: 48.50   Length: 2432
Best Local Similarity: 56.83%   Matches: 113
Query Match: 35.87%   Conservative: 66
                          18.80%   Mismatches: 109
                          4   Gaps: 27
                                         Indels: 11

Qy 154 GluAsnLysSerLeuGluValSerAspThrArgPheHisSerPheTyrGluLeu 173
        |||||||||||||||||||||:|||||||||||||||||||||||||||:::::
Db 1028 GAACAGCACAAATGTTGATATTTGTCAGGGCGATTTGAAGAGAATTCATTCAAACAGTTA 1087

Qy 174 LysAsnValThrAsnAsnPheAspGluArgProIleSerValGlyGlyAsnLysMetGly 193
        ||||||||||||||||||||||||||||||||||||||| :::|||
Db 1088 CGTGTTTCCACTAATAATAATTTTAGTAGCAAG------------------------AATATTTTAGGA 1129

Qy 194 GluGlyGlyPheGlyValValThrTyrLysGlyTyrVal---AsnAsnThrThrValAlaVal 212
        |||||||||||||||||||||||||||||| ::|||||||||||:::
Db 1130 GTAGGAGGATATGGAATTGTCTATAAAAGGATTCCTACAAGATGGGCACTATAGTAGCAATA 1189

Qy 213 LysLysLeuLeuAlaAlaMetValAlaSerIleThrThrGluGluLeuLysGlnIlePheAspAspGln 232
        ::::||||||||||||||||||||||||||||||
Db 1190 AAAAAGGTTGAAA------------------GATGGTAAGTGGGAGGAGGAAAATTCAATTCAAACA 1240

Qy 233 GluIleLysValMetAlaLysCysGlnHisGluAsnLeuValGluLeuLeuGlyPheSer 252
        |||||||||||||||||:::||| |||||:: ||||||:||||||
Db 1241 GAGGTGGAAATGATCAGCTTGCGCGTGCCATAGGAGAACCTATTACGATTGTATGGAGATTTTGC 1300

Qy 253 SerAspSerGlyAspAspLeuCysLeuValThrTyrValTyrMetProAsnGlySerLeuLeuAsp 272
        :::
Db 1301 ACAACCTCCAGAGAGAGCTTCTGGTTCTATCCCATACTACCAAATGGAAGTGTG------ 1354

Qy 273 ArgLeuSerCysLeu--------------AspGlyThrProProLeuSerTrpHisMetArg 288
        :::||| |||||||||||:::
Db 1355 ---GCCTCTGTCTTAGAGATCATAATAATGAAAGCTTGCCCTGGACTGGCCTACTCGC 1411

Qy 289 CysLysIleAlaGlnGlyAlaAlaAsnGlyIleGluAsnPheLeuHisGluAsn--------- 305
        |||||||||||||||||||:: |||:||||||||||||:::
Db 1412 AAGGGTATAGCCCTTGGAGCAGAGCTGTTAATAAATTGCATGGCAAATGTGATCCC 1471

Qy 306 HisHisIleHisIleAlaGlyAsnIleLeuLeuAspAsnGluAlaPheThrAla 325
        ||||||||:||:||||||||||||||||||||||||||
Db 1472 AAGATTATTCACCGGCGATGTGAAAAGCAGCAAATATATTACTGGACATAGTTTGAAGCT 1531

Qy 326 LysIleSerAspPheGlyLeuAlaArgAlaSerGluLysProGluAlaLeu-------- 345
        ::::||||||||||||||||||||||||||||||
Db 1532 GTTGTTGGACAGTTTTGGGCTAGCAAAGCTCGTTGGAT---CACACGGATTCTCATGTGACT 1588

Qy 346 SerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu---ArgGlyVaIIle 364
        ||||||||||||||||||:::|| |||:::
Db 1589 ACTGCTGTGTTCGAGGGACGGTAGGTCACATTCCCCAGAGATACCTTTCAACGGGACAAATCT 1648

RESULT 9
US-09-135-232-1
; Sequence 1, Application US/09135232
; Patent No. 6262228
; GENERAL INFORMATION:
; APPLICANT: Cao, Zhaodan
; TITLE OF INVENTION: IRAK3 Polypeptides, Polynucleotides and Methods
; FILE REFERENCE: T98-019
; CURRENT APPLICATION NUMBER: US/09/135,232
; CURRENT FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2288
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(1851)

US-09-135-232-1

Alignment Scores:
Pred. No.:
Score: 4.1e-41
Percent Similarity: 438.00   Length: 2288
Best Local Similarity: 44.80%   Matches: 127
Query Match: 26.96%   Conservative: 84
                          18.36%   Mismatches: 188
                          3   Gaps: 72
                                         Indels: 12

Qy   16 ValGlyLeuLeuArgLysLeuSerAspPhe IleAspProGlnGluGlyTyrProLysLeu 35
        ::|||||||||||||||||||||||||:: ||||||||||||||||||||||||:::
Db   88 ------------LeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeuLeuProAspPrAla 104

Qy   36 AlaValAlaIleIleLysLysProSerGlyAspAspArgTyrAsnGluPheHisIleArgArg 55
        :::||||||||||||||||||||||||||||||||:::
Db 105 GCTATTCATTTTAATTACAAAACTAT------------------GGAGCAGTGTTGAGTCCTTCAGAG 119

Qy   56 PheGluAlaLeuGlnThrGlyLysSerProThrSerGluLeuPheAspTrpGly 75
        |||||||||||||||||||:||:||||||||
Db 196 GCAGAGAGACTTTCAAGC---------AGCTTGGCTGGGATGTTCGTCAT 234

Qy   76 ThrThrAsnCysThrValGlyAspLeuValAlaAspLeu-------------- 87
        |||||||||:: ||||||||||||||||::
Db 235 ATTGAAAAGTATGTAGACCAAGGTAAAAGTGGAACCAAGAGAAATTACTTTGGTCTCCTGGGCA 294

Qy   96 CAGAAAAACAAGACCATCGGTTGACCATTTTACAGGTCCTCCAGGAGGATGGAGACATCGTCGA 354
        |||||:::
Qy 355 GCTATTCATTTTAATTACAAAACTAT 402
        |||||:::

Qy 105 ValProLysThrAlaAsnThrLeuProSer-------------------LysGluAla 119
        |||::: ||||| |||||

Qy 365 ThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluLeuIleThrGlyLeu 384
        |::: ||||||||||||||||||||||||||||:|||||:::
Db 1649 TCAGAGGAAAACTGATGTATTTTGGCTTTGGGAAGTTGGAACTCATTACAGGACAA 1708

Qy 385 ProLeuAlaValAsp---------GluHisArgGlnProGlnLeuLeuAspIleLysGlu 401
        ||||:|| :::||||||||||||||:: |||
Db 1709 AGGGCTTTAGAGTTTGGCCAGGCTGCCAAACGAAAAAGTTGTAATGCGGATTGGGTAAAG 1768

Qy 402 GluIleGluAspGluGluLysThrIleGluHisProAspLysLysGlyMet---AsnAsp 420
        ||:::| |:::|||| ||
Db 1769 AAGCTTCAT---CAAGAGAACAGAAGTTGCACCTCCTTGCTGATAAGATCTTAAGGACAA 1825

Qy 421 AlaAspSerThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys 440
        |||| :::|||| |||
Db 1826 TTTGATGAGTTGAGCTAGAAAGAGATGGTTCAGGTTTCTTTGGTCGTATGCACCCAATTTCAG 1885

Qy 441 LysAsnLysArgProAspIleLysLysValGlnGlnLeuLeuGln 455
        |||||| |||:: |||:::|||
Db 1886 CCTGGACATCGTCCAAAAAATGTGATGTTTTGAGAATGTTGGAA 1930

```
RESULT 10
US-09-863-549-1
; Sequence 1, Application US/09863549
; Patent No. 6576444
; GENERAL INFORMATION:
;  APPLICANT: Cao, Zhaodan
;  TITLE OF INVENTION: IRAK3 Polypeptides, Polynucleotides and Methods
;  FILE REFERENCE: T98-019
;  CURRENT APPLICATION NUMBER: US/09/863,549
;  CURRENT FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: 09/135,232
;  PRIOR FILING DATE: 1998-08-17
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
;  LENGTH: 2288
;  TYPE: DNA
;  ORGANISM: human
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (64)..(1851)
; DB: US-09-863-549-1

US-10-001-254-16 (1-460) x US-09-863-549-1 (1-2288)

Alignment Scores:
Pred. No.:
Score:                   4.1e-41           Length:         2288
Percent Similarity:      44.80%            Matches:        127
Percent Identity:        26.96%            Conservative:   84
Best Local Similarity:   18.36%            Mismatches:     188
Query Match:             4                 Indels:         72
                                           Gaps:           12
```

Right-column alignment

```
Qy   16 ValGlyLeuIleArgLysLeuSerAspPheIleAspProGlnGluGlyThrLysLysLeu   35
           :::     ||| ||     :::       ||     |:::
Db  136 CTCGGGAGAGCCTCTGCGGTGTTCTGGACAGCTCGCCGACGGCGCGCGCGCGGGCCTG  195

Qy   36 AlaValAlaIleLeuLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArg   55
             ||     ::|    |||            :::        ||
Db  196 GCAGAGAGACTTTCAAGC---------------AGCTGGCGTCGGATGTGTCCTTCAT  234

Qy   56 PheGluAlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGly   75
           ||             ||||     :::      :::       |||
Db  235 ATTGAAAAGTATGTAGCACAAGTAAAAGTGGAACAAGAGAATTACTTGGTCCTGGGCA  294

Qy   76 ThrThrAsnCysThrValGlyValGlyThrThrAla                        87
           ::|           :::       ||
Db  295 CAGAAAAAACAGACCATCGGTGACCTTTTACAGGTCCTCCAGGATGGGACATCGTCGA  354

Qy  105 ValProLysThrAlaAsnThrLeuProSer------------LysGluAla------IleThr 119
           ||                   |||                  |||          :::
Db  355 GCTATTCATTTAATTACAAACTAT--------------------GGAGCAGTGTGTTGAGTCCTTCAGAG  402

Qy  120 ValGlnIleLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGln 139
           |||              :::       |||            :::      |||
Db  403 AAGAGTTATTCAGGAAGGTGGATTTCCAAATATATATTATTCAAGGAAAACAGCCAAGTGCACC 462

Qy  140 AsnLeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGlu 159
           |||              ||||       :::      :::
Db  463 GTGGATAATGTTCTTATTCCTGAACAATAATGAAAAAGGAGTACTGCTT----------- 510

Qy  160 ValSerAspThrArgPheHisSerPheTyrGluLeuLysAsnValThrAsnAsn      179
           :::        |||
Db  511 -------------------------------------------------------AAATCT----- 516

Qy  180 PheAspGluArgProIleSerValGlyGlyAsnLysSerMetGlyValGlyPheGlyVal 199
           |||            :::       |||
Db  517 -------------TCCATCAGCTTTCAAAATATCATAGAAGGAT--------------- 555

Qy  200 ...
Db  556 TTCCACAAAAGACTTCCTA------------------------------------------ 597
```

Left-column alignment

```
Db  403 AAGAGTTATTCAGGAAGGTGGATTTCCAAATATATATTATTCAAGGAAAACAGCCAAGTGCACC 462
Qy  120 ValGlnIleLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGln 139
Db  463 GTGGATAATGTTCTTATTCCTGAACAATAATGAAAAAGGAGTACTGCTT------------ 510
Qy  140 AsnLeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGlu 159
Db  511 -------------------------------------------------------------- 516
Qy  160 ValSerAspThrArgPheHisSerPheTyrGluLeuLysAsnValThrAsnAsn      179
Db  517 -------------TCCATCAGCTTTCAAAATATCATAGAAGGAT--------------- 555
Qy  180 PheAspGluArgProIleSerValGlyGlyAsnLysSerMetGlyValGlyPheGlyVal 199
Db  556 TTCCACAAAAGACTTCCTA------------------------------------------ 597
Qy  200 ValTyrLysGlyTyrValAlaAsnThrThrValAlaValLeuLysLeuAlaAlaMetVal 219
Db  598 GTATACAGAGAGCGGGAGATTCAAAACCTAACATATGCTGTCAAATTATTATTAAACGGAGGAA 657
Qy  220 AspIleThrThrGluGluLysGlnPheAspGlnLeuLeuLysGlyValMetAlaLys 239
Db  658 AAAATGCAGTGTAAGAAGCAATGGAAGAGGTTTTTATCTGAGCTTGAAGTTTTACTACTG 717
Qy  240 CysGlnHisGluAsnLeuValGluLeuValGlyPheSerSerAspGlyAspAspLeuCys 259
Db  718 TTTCATCACCCAAACCATTACTAGAGTTCGGCTGCATATTTACAGAGACTGAGAAGTTCTGT 777
Qy  260 LeuValTyrArgTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGly 279
Db  778 CTGATTTATCCATACATGAGAAATGGAACACTTTTGCACAGATTGCAGTGTGTAGGTGAC 837
Qy  280 ThrProProLeuSerTrpHisMetArgCysLysLysIleLeuAlaGlnAlaAlaAsnGlyIle 299
Db  838 ACGGCCCACTCCCTTGGCACATTCGAATCGGTATAATTAATTAGGAATATCAAAAGCATT 897
Qy  300 AsnPheHisGluHisGluAsnHisHis------                  IleHisArgSerAlaSerAlaAsn 316
Db  898 CACTACCTGCACAACGTTCAACCATGGTCGTCTGTCCAGTATTAATTAACAGAATGTGTTCTAC 957
Qy  317 IleLeuLeuAspGluAlaPheThrAlaLysIleIleSerGluAspPheGlyLeuAlaArgAlaSer 336
Db  958 ATCCTTTTGGATGATGCAGTTTCAACCCAACTAACTAACTTTGCCATGGCACACTTCCGG 1017
Qy  337 GluLysGlyPheAlaGlnThrHisValMetThrSerArgIleValGlyThrThrArgThrAla 353
Db 1018 TCCCACCTAGGACACATCAGAGTTGTACCATAAATATGACCAGCAGTGTAAAACATCTG 1077
Qy  354 --TyrMetAlaProGluAlaLeuArg--GlyGluIleThrProLysSerAspIleTyr 371
Db 1078 TGGTACAATGCCAGAGAGTACATCAGACAGGGAAACTTCCATTAAAACAGAGTGTTCTAC 1137
Qy  372 SerPheGlyValValLeuLeuGluLeuIleThrGlyLeuProAlaValAspGluHisArg 391
Db 1138 AGCTTTGGAAGTTGTAATAGAAGTTCTACAGGAGTAGAGTGTTAGATGNTCCA 1197
Qy  392 GluProGlnLeuLeuLeuAspLeuLysGluGluLeuGluGlu------LysThr 409
Db 1198 AAACAATATCAGCTCGCGGGACTCTCTTAGAGAATTGATGGAGAAGAGGCCTGGAATTCA 1257
Qy  410 IleGluAspTyrIleLeuAspLysLysMetAlaAsnAspAlaAspSerThrSerValGluAlaMet 429
Db 1258 TGTCTCTCATTTCTAGATAAGAAATAGGCCTCCCGGAATTTCCTGCCAAGCTC 1317
Qy  430 TyrSerValAlaSerGlnCysLeuHisGluLysLysAsnLysGlyArgProAspIleLysLys 449
Db 1318 TTCTGTTTGGCAGGCCCCGTGTGCCTGCAAGCGCGGGCAAAGTTAAGACCCATCAATGGATGGATGAA 1377
Qy  450 ValGlnGlnLeuLeuGlnGluMetThrAlaSer 460
Db 1378 GTTTTAAATACTCTTGAAAGTACTCAAGCCAGC 1410
```

```
;
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO: 4
; LENGTH: 2868
```

RESULT 11
US-09-228-986-4
Sequence 4, Application US/09228986
Patent No. 6359198

```
;
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-228-986-4

Alignment Scores:
Pred. No.:              3e-40
Score:                  432.00
Percent Similarity:     55.35%
Best Local Similarity:  35.53%
Query Match:            18.11%
DB:                     4

              Length:      2868
              Matches:     113
              Conservative: 63
              Mismatches:  112
              Indels:      30
              Gaps:        12
```

US-10-001-254-16 (1-460) x US-09-228-986-4 (1-2868)

Qy 200 ValTyrLysGlyTyrValAsnAsnThrThrValAlaValLysGlyLeuAlaAlaMetAlaVal 219
         |||| ::::|| |||| :::::
Db 598 GTATACAGAGGTGAGAGATTCAAAAACCTAACATATGTGTCAAATTATTAAACAGGAGAA 657

Qy 220 AspIleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysValMetAlaLys 239
Db 658 AAAATGCAGTGTAGAAGCATTGGAAGAGGTTTTTATCTGAGCTTGAAGTTTACTACTG 717

Qy 240 CysGlnHisGluAsnLeuValGlyLeuLeuGlyPheSerSerAspGlyAspLeuCys 259
Db 718 TTTCATCACCCAAACATACTAGAGTTGGCTGCATATTTTACAGAGACTGAGAGTTCTGT 777

Qy 260 LeuValTyrValTyrMetProAsnGlySerLeuLeuSerCysLeuAspGly 279
Db 778 CTGATTTATCCATACAATGAGAAATGGAACACTTTTTTGACAGATTCCAGTGTGTAGGTGAC 837

Qy 280 ThrProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIle 299
Db 838 ACGGGCCCACTCCCTTGGCACATTCGAATCGGTATATTAATTAGGAATATTCCAAAGCCATT 897

Qy 300 AsnPheLeuHisGluAsnHisHis---------IleHisArgAspIleLysSerAlaAsn 316
Db 898 CACTACCTGCCACACGTTCAACCATGCTCGGTCATCTGTGGCCAGTATATCAAGTGCAAAC 957

Qy 317 IleLeuLeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSer 336
Db 958 ATCCTTTTGATGACAGTTTCAACCAAACTAACTGATTTTGCCATGGGCACACTTCCGG 1017

Qy 337 GluLysPheAlaGlnThrValMetThrSerArgIleValGlyThrThrAla------ 353
Db 1018 TCCCACCTAGAACAATCAGAGTTGTACCCATAAAAATGACCAGCAGCAGTAAACATCTG 1077

Qy 354 ---TyrMetAlaProGluAlaLeuArg---GlyGluIleThrProLysSerAspIleTyr 371
Db 1078 TGGGTACCAATGCCAGAAGAGTACATCAGACGGGGAAACTTTCCATTAAAACAGATGCTAC 1137

Qy 372 SerPheGlyValValLeuLeuGluIleIleThrGlyLeuProAlaValAspGluHisArg 391
Db 1138 AGCTTTTGGAATTGTGATAAATGGAAGTTCTACAGGAGTAGTGTTAGATGATCCA 1197

Qy 392 GluProGlnLeuLeuLeuAspProIleLysGlyLysIleGluAspGluGlu------LysThr 409
Db 1198 AAACATATCCAGCTCGGGATCTCCTTAGAGAAATTGATGGAAGAGAGGCCTGGATTCA 1257

Qy 410 IleGluAspThrIleLeuAspProLysMetAsnAspAlaAlaSerThrSerValGluAlaMet 429
Db 1258 TGTCTCTCATTTCTAGATAAGAAAGTGCCTCCCCTCCGGAATTTCTCTGCCAAGCTC 1317

Qy 430 TyrSerValAlaSerGlnCysLeuHisIleGlyLysLysAsnLysArgProAspIleLysLys 449
Db 1318 TTCTCGTTTGCCAGGGCGGTGTGCTCGAACGGGGAAAGTTAAGACCATCAATGGATGAA 1377

Qy 450 ValGlnGlnLeuLeuGlnLeuGlnMetThrAlaSer 460
Db 1378 GTTTTAAATACTTCTTGAAAGTACTCAAGCCAGC 1410

US-10-001-254-16 (1-460) x US-09-228-986-4 (1-2868)

Qy 153 ProGluAsnLysSerLeuLeuValSerAspThrArgPheHisSerPheTyrGlu 172
Db 1071 CCTGCTGAGGAGGATCCAGAAGTTCACTTAGCCCAAACTTAAGAGAGGTTCTCAATTACGGGAA 1130

Qy 173 LeuLysAsnValThrAsnAsnPheAspGluArgProIleSerValGlyGlyAsnLysMet 192
Db 1131 TTACAGGTTGCAAACTGATAGGTTTTTAGACAATAGA---------AACATTCTT 1172

Qy 193 GlyGluGlyGlyPheGlyValValLysGly---TyrValAsnAsnThrThrValAla 211
Db 1173 GGCAGAGGTGGTTTTGGAAAAGGTGTACAAAAGGGGCGGCCTTGCAGATGGTTCCTTGGTGGCT 1232

Qy 212 ValLysLysLeuAlaAlaMetAlaValAspIleThrThrGluGluLeuLysGlnGlnPheAsp 231
Db 1233 GTAAAAACGTCTGAAG---------GAAGGACGTACACCGGGTGGGACAGTTCCAGTTTCAA 1283

Qy 232 GlnGluIleLysValMetAlaLysCysGlnHisGluAsnLeuValGlyLeuLeuGlyPhe 251
Db 1284 ACAGAAGTGGAGATGAATAAGCATGGCAGTACAGAACCTTCCTTCGACTACGTGGATTC 1343

Qy 252 SerSerAspGlyAspLeuCysLeuValTyrValTyrMetProAsnGlySerLeuLeu 271
Db 1344 TGCATGACACCCACTGAACGGCTGCTTGTTTATCCCTACATGGCCAATGGAAGTGTT--- 1400

Qy 272 AspArgGluSerCysLeu---------AspGlyThrProProLeuSerTrpHisMet 287
Db 1401 ------GCTTCATTGCCTACAGAGAGGGCACAAAAATGACCCAGCTTAGATTGGCCAACT 1454

Qy 288 ArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleIleAsnPheLeuHisHisAsn------ 305
Db 1455 CGCAACGGCCATAGCATTGGGGTTCTCCAAGAGGTCTCTTCCACTTGCATGATCATTGTGAT 1514

Qy 306 ---HisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAspGluAlaPheThr 324
Db 1515 CCTAAGATTATTCACCGGGATGTCAAGGCTCTAACGATCTTACTGGATGAAGAATATGAG 1574

Qy 325 AlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAlaGlnThrValMet 344
Db 1575 GCAGTGTTGGGGGGATTTTGGCTTGCCAAAACTTTATGGAT---TATAAGGACACAATGTT 1631

Qy 345 ThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu---ArgGlyGlu 363
Db 1632 ACGAGCGGCTGTTCGGGGAACCCATTGGCCACCTGAGTACTTACTTGGAAAG 1691

Qy 364 IleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIleIleThrGly 383
Db 1692 TCTTCCGAAAAGACAGACGTATTTGGATATGGAAACATCAGTTACGGGA 1751

Qy 384 LeuProAlaValAspGluHisArg---------GluProGlnLeuLeuLeuAsp--- 398
Db 1752 CAACGGCGCATTTGACCTTGCACGTTTAGCAAATGATGATGAGATGGTTGCTTGGACTGG 1811

Qy 399 IleLeuGlyGluIleGluAspGluGluLysThrIleGluAspThrIleLeuAspProLysMet 418
Db 1812 GTTAAAGGCCTACTA---AAAGAGAAGAGGCTTGAATATGGTAGTTAGTTGAATCCGATCTT 1865

Qy 419 ---AsnAspAlaAspSerThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeu 437
Db 1866 AAGAACAAATTATGTTGAAGCAGAGAGGTGGACAACTTATTATTCAAGTTGCATTACTTGTACA 1925
```

```
Qy   438 HisGlyLeuGlyLysAsnLysArgProAspIleLysLysValGlnGlnGlnLeuLeuGln 455
             |||:||     |:||:::      ::|:::||:    :||::::|||::
Db  1926 CAAGGGTCACCAATGGATAAGACCAAAGATGTCTGAAGTGGTAAGGATGTTGGAA 1979
```

RESULT 12

US-08-447-185-2

Sequence 2, Application US/08447185

Patent No. 5648599

GENERAL INFORMATION:

  APPLICANT: Tanksley, Steven D.

  APPLICANT: Martin, Gregory B.

  TITLE OF INVENTION: GENE CONFERRING DISEASE RESISTANCE

  TITLE OF INVENTION: TO PLANTS BY RESPONDING TO AN AVIRULENCE GENE IN PLANT PATHOGE

  NUMBER OF SEQUENCES: 5

  CORRESPONDENCE ADDRESS:

  ADDRESSEE: Michael L. Goldman

  STREET: Clinton Square, P.O. Box 1051

  CITY: Rochester

  STATE: New York

  COUNTRY: U.S.A.

  ZIP: 14603

  COMPUTER READABLE FORM:

  MEDIUM TYPE: Floppy disk

  COMPUTER: IBM PC compatible

  OPERATING SYSTEM: PC-DOS/MS-DOS

  SOFTWARE: PatentIn Release #1.0, Version #1.25

  CURRENT APPLICATION DATA:

  APPLICATION NUMBER: US/08/447,185

  FILING DATE:

  CLASSIFICATION: 800

  PRIOR APPLICATION DATA:

  APPLICATION NUMBER: US/08/111,078

  FILING DATE:

  ATTORNEY/AGENT INFORMATION:

  NAME: Goldman Mr., Michael L.

  TELECOMMUNICATION INFORMATION:

  TELEPHONE: (716)263-1000

  TELEFAX: (716)-263-1600

  INFORMATION FOR SEQ ID NO: 2:

  SEQUENCE CHARACTERISTICS:

  LENGTH: 966 base pairs

  TYPE: nucleic acid

  STRANDEDNESS: double

  TOPOLOGY: unknown

  MOLECULE TYPE: DNA (genomic)

US-08-447-185-2

US-10-001-254-16 (1-460) x US-08-447-185-2 (1-966)

Alignment Scores:

Pred. No.:	
Score:	7.69e-41
Score:	430.50
Percent Similarity:	49.55%
Best Local Similarity:	34.03%
Query Match:	18.04%
DB:	1

Length:	966	
Matches:	114	
Conservative:	52	
Mismatches:	100	
Indels:	69	
Gaps:	12	

```
Qy   161 SerAspThrArgPheHisSerPheSerTyrGluLeuLysAsnValThrAsnAsnPhe 180
             |:|::|||:|:|     ||:||:::|     ||::|:||::|:
Db    79 --------------TATCGAGTTCCTTAGTAGTAGTTGAGGAGGCAACTAATAATTTT 123

Qy   181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal 200
             ||:||||:       :::||          |||:|:||:||||||:
Db   124 GATCACAAGTTTTA----------------ATTGGACATGGTGTCTTTGGGAAGGTT 165

Qy   201 TyrLysGlyTyrVal---AsnAsnThrThrValAlaValLysLeuAlaAlaMetVal 219
             |||||||:|||||         ||   |||:|:|::|||:::||:|:
Db   166 TACAAGGGTGTTTTGCCTGATGGAGCAAAAGTGGGCCCTGAAAAGG---------- 210

Qy   141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal 160
             ||||||       :||::||     |||:::||||:||
Db    46 TTAAGCTCGAGTTATCTCGTTCCTTTGAAAGT-------------------------- 78
```

```
Qy   220 AspIleThrThrGluGluGluLeuLysGln---------GlnPheAspGlnGluIleLysVal 236
             |||          ::::|:|::|::        |::|::||:|:||:||:|
Db   211 ------CGTACACCTGAGTCCTCCACAGGTATTGAGAGTTGAAACAGAAAATTGAGACT 264

Qy   237 MetAlaLysCysGlnHisGluAsnLeuValGluLeuGlyTyrProAlaIleLysSerAla 256
             |::|     ::||:|   |||:||:||     |||:   :|||::|||::
Db   265 CTCTCATTTTGCAGAACATCCGCATCTCGGTTTCATTGAATTGATAGGATTTTCGTGATGAAAGAAAT 324

Qy   257 AspLeuCysLeuValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCys 276
             |||:|||:||:   ||:||:|:||     |:|||||||:|:|:|::||:
Db   325 GAGATGAATTCTAATTTATTAAAATACATGGAGAATGGGAACCTCAAGAGACATTTGTATCGA 384

Qy   277 LeuAsp---GlyThrProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAla 295
             ||:|:        :|||  ||||:|:||::|:||||:::||:|:|:
Db   385 TCAGATCTACCCACCAATGAGCATGGGGAGCAGAGGCTGGAGATATGCATAGGGGCA 444

Qy   296 AlaAsnGlyIleAsnPheLeuHisGluAsnHisHisIleHisHisArgLysSerArgAla 315
             |::|     :||||     |||::||:||::||     ||:||:|::||
Db   445 GCCAGAGGTCTACACTACCTTCATACTAGAGACAAAATTATACATCGTAGAGTCTATA 504

Qy   316 AsnIleLeuLeuAspGluAlaPheThrAlaLysIleSerGluAspPheGlyLeuAlaArgAla 335
             |:|||:|||::|::|:|     |||     ||:|||:||     ||:|:
Db   505 AACATATTGCTTGATGAGAATTTTGTGCCAAAAAATTACTGATTTTGGAATTGCAAGAA 564

Qy   336 SerGluLysPheAlaGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMet 355
             ::|     |:||||||  :||  |||:|::|||||:|:|:||   :|::
Db   565 GGGACTGAGCTTGATCAAAACCCATCTTAGCACAGTAGTGAAAGGACAACTCTCGGCTACATT 624

Qy   356 AlaProGlu---AlaLeuArgGlyGlyIleThrProLysSerAspIleTyrSerPheGly 374
             |:|:|        |||  :||::     ||:|||     ||::||:||:
Db   625 GACCCTGAAATATTTATTATAAAGGGACGACTCACTGAAAAAAATCTGAATGTTTATTCTTCGGT 684

Qy   375 ValValLeuLeuGlu-------IleIleThrGlyLeuPro------ 385
             ||||        |             ||:|||:|||
Db   685 GTTGTTTTATTGGAAGTTCTTTGTGCTAGGTGTGCCATAGTTCAATCTCTTCCAAGGGAG 744

Qy   386 -----------AlaValAspGluHisArgGluProGlnLeuLeuLeuAsp 398
                        ||||:|||::|     |||:|:||::|:
Db   745 ATGGTTAATTAAGCTGAATGGGCCAGTGGAGTCGCCATAATAATGGACAGTTG-- 795

Qy   399 IleLysGluGluIleGluLysThrIleGluAspTyrIleAspLysLysMet 418
             |||:|:|||     ||:|:||||:|:||::|
Db   796 ------GAACAAAATCGTAGATCCC----------------------- 813

Qy   419 AsnAspAlaAspSerThrSerValGluAlaMetTyr-------SerValAlaSerGln 435
             |:|:|:|||||     |||  ::|:::|      |::
Db   814 AATCTTCGAGATAAAAATAAGACCAGAGACTCCCTCAGGAAGTTTGGAGATACACGGTAAA 873

Qy   436 CysLeuHisGluLysLysAsnLysArgProAspIleLysLysVal 450
             ||:||     :||:||:|||:|:|:||   |||:|:||||:
Db   874 TGCTTAGCTTTGTCTAGTGAGGATAGGCCATCAAATGGGTGATGTG 918
```

RESULT 13

US-08-447-185-3

Sequence 3, Application US/08447185

Patent No. 5648599

GENERAL INFORMATION:

  APPLICANT: Tanksley, Steven D.

  APPLICANT: Martin, Gregory B.

  TITLE OF INVENTION: GENE CONFERRING DISEASE RESISTANCE

  TITLE OF INVENTION: TO PLANTS BY RESPONDING TO AN AVIRULENCE GENE IN PLANT PATHOGE

  NUMBER OF SEQUENCES: 5

  CORRESPONDENCE ADDRESS:

  ADDRESSEE: Michael L. Goldman

  STREET: Clinton Square, P.O. Box 1051

  CITY: Rochester

  STATE: New York

  COUNTRY: U.S.A.

  ZIP: 14603

  COMPUTER READABLE FORM:

  MEDIUM TYPE: Floppy disk

  COMPUTER: IBM PC compatible

```
;  OPERATING SYSTEM:  PC-DOS/MS-DOS
;  SOFTWARE:  Patentin Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER:  US/08/447,185
;  CLASSIFICATION:
;  FILING DATE:
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER:  US/08/111,078
;  FILING DATE:
;  ATTORNEY/AGENT INFORMATION:
;  NAME:  Goldman Mr., Michael L.
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE:  (716) 263-1000
;  TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO: 3:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH:  2443 base pairs
;  TYPE:  nucleic acid
;  STRANDEDNESS:  double
;  TOPOLOGY:  unknown
;  MOLECULE TYPE:  DNA (genomic)
US-08-447-185-3

Alignment Scores:
Pred. No.:
Score:                        3.46e-40
Percent Similarity:           430.50
Percent Similarity:           49.55%
Best Local Similarity:        34.03%
Query Match:                  18.04%
DB:                           1

Length:    2443
Matches:   114
Conservative:  52
Mismatches:    100
Indels:        69
Gaps:          12

US-10-001-254-16 (1-460) x US-08-447-185-3 (1-2443)

Qy  141  LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal 160
         ||||                ||||                          |
Db  109  TTAAGCTCGAGTTATTCGTTCCTTTTGAAAGT------------------------ 141

Qy  161  SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe 180
              :::             ||||||   |:::::          ||||     :
Db  142  ------------------TATCGAGTTCCTTTAGTAGATTTGGAGGAGCAACTAAATTTT 186

Qy  181  AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal 200
         |||      :::                                             ||
Db  187  GATCACAAGTTTTA---------------------ATTGGACATGGTCTCTTTGGAAGGTT 228

Qy  201  TyrLysGlyTyrTyrVal---AsnAsnThrThrValAlaValAlaValLysMetVal 219
         |||||||||          ::               ||||||
Db  229  TACAAGGGGTGTTTTGCGTGATGGAGCAAAGGTGGCCCCTGAAAAGG------------ 273

Qy  220  AspIleThrThrGluGluLeuLysGln------GlnPheAspGlnGluIleLysVal 236
              :                    :::|||||||||||:::|||
Db  274  ------CGTACACTGAGTCCTCACAAGGTATTGAAGAGTTCGAAACAGAAATTGAGACT 327

Qy  237  MetAlaLysCysGlnHisGluAsnLeuValGluLeuGlyPheSerSerSerGlyAsp 256
         |||    :::
Db  328  CTCTCATTTTGCAGACATCCGCATCTGGTTCATTGATTCGTGATGAAAGAAAT 387

Qy  257  AspLeuCysSerLeuValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCys 276
         :::::|||||||:::|||              ||||||::::|||
Db  388  GAGATGAATTCTAATTTATAAATACATGGAGAAATGGGGAACCTCAAGACACATTGTATGGA 447

Qy  277  LeuAsp---GlyThrProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAla 295
            |||            ||||||                      |||
Db  448  TCAGATCTACCCACAAATGAGCATGAGCTCGGAGCAGCAGAGGCTGGACATATCCATACGGGGCA 507

Qy  296  AlaAsnGlyIleAsnPheLeuHisGluAsnHisHisIleHisArgAspPheLysSerAla 315
         |||||||||||||||||
Db  508  GCCAAGAGGTCTACACTACCTTCATACTAGAGCAATTATCATCGTTAGTCAAGTCTATA 567

Qy  316  AsnIleLeuLeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAla 335
         |||||||||||||||                     ||||
Db  568  AACATATTGCTTGATGAGAATTTTGTGCCAAAAAATTACTGATTTTGGAATATCCAAGAAA 627
```

```
RESULT 14
US-07-717-331F-4
;  Sequence 4, Application US/07717331F
;  Patent No. 5484905
;  GENERAL INFORMATION:
;  APPLICANT:  June Nasrallah; Michael Nasrallah; and Joshua
;  APPLICANT:  Stein
;  TITLE OF INVENTION:  A Receptor Protein Kinase Gene
;  TITLE OF INVENTION:  Encoded At The Self-Incompatability Locus
;  NUMBER OF SEQUENCES:  9
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE:  Yahwak & Associates
;  STREET:  25 Skytop Drive
;  CITY:  Trumbull
;  STATE:  Connecticut
;  COUNTRY:  USA
;  ZIP:  06611
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE:  Floppy Disk
;  COMPUTER:  Macintosh
;  OPERATING SYSTEM:  MS-DOS
;  SOFTWARE:  Microsoft Word
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER:  US/07/717,331F
;  FILING DATE:  June 19th 1991
;  CLASSIFICATION:  800
;  ATTORNEY/AGENT INFORMATION:
;  NAME:  George M. Yahwak
;  REGISTRATION NUMBER:  26,824
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE:  (203) 268-1951
;  TELEFAX:  (203) 268-1951
;  INFORMATION FOR SEQ ID NO: 4:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH:  2749 base pairs
;  TYPE:  nucleic acid
;  STRANDEDNESS:  single
;  TOPOLOGY:  linear
;  MOLECULE TYPE:  DNA
US-07-717-331F-4

Alignment Scores:
Pred. No.:
Score:                        9.43e-40
Percent Similarity:           427.50
Percent Similarity:           44.25%

Length:    2749
Matches:   152
Conservative:  75
```

```
Qy  336  SerGluLysPheAlaGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMet 355
         ||||||   ::              ::::::::                ||||||  :::
Db  628  GGGACTGAGCTTGATCAAACCCATCTTAGCACAGTAGTGAAAGGAACTCTCGGCGTACATT 687

Qy  356  AlaProGlu---AlaLeuArgGlyGluIleThrProLysSerAspIleTyrSerPheGly 374
         ||||||           :::|||||||||::||||||||:::||::::||||||||||
Db  688  GACCCTGAAATATTTTATAAAGGGAGGACGAACTCACTGAAAAATCTGATGTTAATTCTTCGGT 747

Qy  375  ValValIleLeuGluGlu--------------------IleIleThrGlyLeuPro---- 385
         |||           :::                          :::
Db  748  GTTGTTTTTTATTCGAAGTTCTTTGTCTAGGTCTCCATAGTTCAAATCTTCCAAGGGAG 807

Qy  386  -----------AlaValAspGluHisArgGluProGlnLeuLeuAsp 398
                    :::        ||||      :::||||||||
Db  808  ATGGTTAATTTAGCTGAATGGGCAGTGAGGTCGCATAATAATGGACACAGTTG-------- 858

Qy  399  IleLysGluGluIleGluAspGluGluLysThrArgIleGluAspTyrIleAspLysLysMet 418
               |||:::|||         |||||||  :::|||
Db  859  ------GAACAAATCGTAGATCCC--------------------------------------- 876

Qy  419  AsnAspAlaAspSerThrSerValGluAlaMetTyr-----------SerValAlaSerGln 435
         |||                 |||||::                        |||
Db  877  AATCTTGCAGATAAAAATAAGACCAGAGTCCCTCAGGAAGTTTGGACATACAGCGGTAAA 936

Qy  436  CysLeuHisGluLysGlyLysAsnLysArgProAspIleLysLysVal 450
         ||||||                ||   :::                   |||
Db  937  TGCTTAGCTTTGTCTAGTGAAGATAGGCCATCAATCGGGTGATGTG 981
```

Best Local Similarity: 29.63%
Query Match: 17.92%
DB: 1

Mismatches: 157
Indels: 129
Gaps: 24

US-10-001-254-16 (1-460) x US-07-717-331F-4 (1-2749)

```
Qy    1 MetAsnLysProIleThrProSerThrTyrVal---ArgCysLeuAsnValGlyLeuIle 19
          |||||                     :::        ::::|||||||||||
Db 1099 ATGAATTTGCCGGATACTAAGAGACGGCAACGTGGGATCGGACAATGGAAAAAATGT 1158

Qy   20 ------------ArgLysLeuSerAsp------------PheIleAspProGlnGlu 30
                  |||||||||                   :::
Db 1159 GAAGAGAGGGGTGCTTAGCGATGTGTAACTGTACTTCGTTTGCTATTGCGGATGTTCGGAAT 1218

Qy   31 Gly---------------------TrpLysLeuLeuAlaValAlaIleLysLys----Pro 42
         |||                       :::|||||||||
Db 1219 GGAGGATTGGGGTTGTGTTGTGTTTGGACCCGGAGACTCCGTTGCGATCKCAGGAAATTCGCCGTC 1278

Qy   43 SerGlyAspAspArgTyrAsnGlnPheHis------------IleArgArgPheGlu 57
         |||||||||||||             :::            |||
Db 1279 GGTGGTCAAGATCTTAGCTGCAGATTGAATGCTGCTGACTTAGATATTTCCGAAAATTCGGAG 1338

Qy   58 AlaLeuLeuGlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThr 77
         ::::|||||||||                      |||
Db 1339 AAGAGACACCGAACTGGAAA------------------ATCATAGGTTGG------------ 1371

Qy   78 AsnCysThrValGlyAspLeuValAlaAspLeuLeuIleGlnAsnGluPhePheAlaProAla 97
             :::|||                      ::::
Db 1372 ------AGTATTGGATCCAGCGGTTATCGTTATTCTG------------ 1401

Qy   98 'SerLeuLeuLeu------------------------------ProAspAlaVal 105
         |||                                       |||
Db 1402 AGTGTTATTCTTGTTCTCGCTTTCTTGGAGGAGGAACAAAAGCAACGCAGATGCAACA 1461

Qy  106 ProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrValGlnGlnLysGlnMet 125
         |||                                        :::|||
Db 1462 CCTATTGTGGGAAATCAAGTCTTAATCAAACGAG--GTGGTGTTACCAAGAAAGAGAGA 1518

Qy  126 ProPheCysAspLysAspArgThrLeuMetThrProValGlnAsnLeuGlnSerTyr 145
                |||                              :::
Db 1519 AATTTTTCTGGAGGAC------------------GATGTAGAAAATTTGGAACTTCCATTG 1563

Qy  146 MetProProAspSerSerGluProGluAsnLysSerLeuGluValSerAspThrArgPhe 165
         |||
Db 1564 ATG-------------------------------------------------------- 1566

Qy  166 HisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPheAspGluArgProIle 185
         |||                               :::
Db 1567 ----------TTCATGAAACGAAGTTAGGCTAATGCAAAGCTTCAGCGACAATAAT 1608

Qy  186 SerValGlyGlyAsnLysMetGlyGlyGlyPheGlyValValTyrLysGly---Tyr 204
         |||                     :::|||
Db 1609 ------------TTTAACAAGGTCCGGAAAAGTGGTTTGGTGTTTTACAAGGAAGGTTA 1659

Qy  205 ValAsnAsnThrThrValAlaLysValAlaMetValAspProIleThrThrGlu 224
         |||||||||||               :::|||
Db 1660 GTTGACCGGCAAGAAAATTGCACGTGAAGAGACTATACGGAAAATGTCAGCTCAAGGTACCGAT 1719

Qy  225 GluLeuLysGlnIleGlnArgGluLeuLysValAlaLysCysSerGlnHisGluAsn 244
         |||||||||||                    :::
Db 1720 GAG-------------------TTCATGAAACGAAGTTAGGCTAATGCAAAGCTTCAGCGACAATAAT 1767

Qy  245 LeuValGluLeuLeuGlyPheSerSerAspGlyAspLeuCysSerLeuValTyrValTyr 264
         ::::|||                       |||
Db 1768 CTTGTCCGACTTCTTGGCCTGTTGTTGTTTATGAGGGCGGAGGATCTTAATTTACGAGTAC 1827

Qy  265 MetProAsnGlySerLeuLeuAspArgLeu------------SerCysLeuAspGly 279
         |||||||                      ::::|||
Db 1828 TTGGAGAATCTAAGCCTCGATTCTCATCTCTTTGATGAAACCAGAAGCTGTATG----- 1881

Qy  280 ThrProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIle 299
         |||                            :::
Db 1882 ----------TTAAATTGGCAAATGAGATTTGATATTATCAATGGTAATTGCCCGAGGGCTT 1932

Qy  300 AsnPheLeuHisGluAsnHisHis-------------IleHisArgAspIleLysSerAlaAsn 316
         ::::||||||||                        |||||||
Db 1933 CTCTATCTTCACCAAGATTCACGGTTTAGAAATCATCACACAGGGATTTGGAAAGCAAGCAAT 1992

Qy  317 IleLeuLeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSer 336
         ::::||||||                      |||
Db 1993 GTCCTTGCTTGATAAAGATATGACTCCAAAAAATTTCAGACTTTGGAATGGCCTAGGATCTTT 2052

Qy  337 GluLysPheAlaGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAla 356
         |||                             :::|||
Db 2053 GGACCGGGATGAGGACGGGAAGCTGACACGGAGGAAGGTGGTCGGGAACTTATGGGCTACATGTCT 2112

Qy  357 ProGlu---AlaLeuArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyVal 375
         |||                              :::|||
Db 2113 CCAGAAATATGCGATGAACGGGGACATTCTCAATGAAAGTCAGAGTGTTTCAGTTTGGGTC 2172

Qy  376 ValLeuLeuGluIleIleThrGlyLeuProAlaValAlaAspGluHisArgGluProGluLeu 395
         ::::||||||                      :::
Db 2173 TTGCTTCTTGAAATTATAAGTGGC-----------AAGAGGAACAAAAGGC 2211

Qy  396 LeuLeuLeuAspIleIleLysGluGluIleGlu--------------AspGlu 406
         |||                                         :::
Db 2212 TTATGCGACTCCGGATAGTGACCCTTAATCTTCTCGGATGGTGATGGGAAATTGGAAAGAGGA 2271

Qy  407 GluLysThrIleGluAspTyrIleAspLysLysMetAsnAspAlaAspSerThrSer---- 425
         |||                                     :::
Db 2272 GGTCAAGGTCTAGAG---ATAGTAGCAAAGGTCATCATAGATTCTTCATCACCAACGTTC 2328

Qy  426 -----------ValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLysLys 441
                    :::
Db 2329 AGCCCACGTGAAATCTTAAGATGCTTACAAATTGGCCTCTTGTCGTTCAAGAACGTCTG 2388

Qy  442 AsnLysArgProAspIleLeuLysValGlnGlnLysLeuGlu 454
         |||                            :::
Db 2389 GAGGATAGACCAATGATCTCGTCAGTAGTTTTTGATGCTC 2427
```

```
LENGTH: 3459 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
    NAME/KEY: CDS
    LOCATION: 34..1908
US-08-980-060-3

Alignment Scores:
Pred. No.:
Score:                  3.5e-39
Percent Similarity:     424.00
Best Local Similarity:  46.19%
Query Match:            25.64%
                        17.77%
DB:                     2

US-10-001-254-16 (1-460) x US-08-980-060-3 (1-3459)

                                  Length:       3459
                                  Matches:      121
                                  Conservative: 97
                                  Mismatches:   172
                                  Indels:       82
                                  Gaps:         14
```

```
Qy   53 IleArgArgPheGluAlaLeuLeuGlnThr-----GlyLysSerProThrSerGluLeuLeu  71
           |||                :::               ::|                |||
Db  157 CTGCGGAAGATCAAGTCCATGGAGCGCGGGTGCCAGGGTGTGACATCACGGCGGGAGCTGCTG  216

Qy   72 PheAspTrpGlyThrThrAsnCysThrValGlyAlaAspLeuValLeuIleGlnAsn  91
           |||||||               |||||||                 :::
Db  217 TGGTGGTGGGGCATGCGGGAGGCCACCGTCCAGCAACTTGTGGACTTCCTGTGCCGCCTG  276

Qy   92 GluPhePheAlaProAlaSerLeuLeuLeu  101
           |||                  :::||||
Db  277 GAGTTCTACCGGGCTGCCCAGATCATCCTGAACTGGAAACCGGCTCCTGAAATCAGGTGT  336

Qy  102 -------ProAspAlaVal---ProLysThrAlaAsnThrLeuProSerLys  115
               |||             :::
Db  337 CCCATTCCAGCCTTCCCTGACTCTGGTGAAGCCAGAAAAAGCCTTTGGCAGCTTCTGTAAGA  396

Qy  116 GluAlaIleThrValGlyHisGlnLysGlnMetProPheCysAspGlyAspArgThrLeuMet  135
           :::||||                |||
Db  397 AAGGCTGAGGATGAACAGGAAGAGGGGCCAGCCTGTG-----AGGATGGCCACCTTTCCA  450

Qy  136 ThrProValGlnAsnLeuGluGlnSerTyr------MetProProAspSer  150
           |||                        |||||||
Db  451 GGCCCAGGGGTCCTCTCCAGCCAGCCCCACCACCCGGCCCTTTCTCCAGCCTCCTGAAGAA  510

Qy  151 SerSerProGluAsnLysSerLeuGluVal----------SerAspThrArgPheHisSer  167
           ::|||            ::               :::
Db  511 GATGCGCCCTCATTCCTTGACGAAGCGACCTCCCCACTTCGTCTGATTCAAAGGACTTCAGC  570

Qy  168 --------------------------------PheSer  169
                                           :::||
Db  571 ACCTCCATTCCTAGCCAGGAAAAAACTTTGAGCTTGGCTGGAGACAGCCTTTCTGGAGT  630

Qy  170 PheTyrGluLeuLysAsnValThrAsnAsnPheAspGluArgProIleSerValGlyGly  189
           |||||||||            :::::::
Db  631 GAGGCAGACCTGGTCCAGGGCAACCGATGACTTCAATCA-----  672

Qy  190 AsnLysMetGlyGluGlyPheGlyValValTyrValLysTyrTyrValAsnAsnThrThr  209
           ::|||:::         :::
Db  673 CGCAAAATCAGCCAGGGAGACCTTTGCTGCTACTACAGGCGCACGGGGAAGCCA  732

Qy  210 ValAlaValLysLysLeuLeuAlaAlaMetValAspIleThrThrGluLeuLeuLysGlnGln  229
           |||||||||            :::
Db  733 TTCGTCTTCAAGAAGCTTCAGAGAGCAGGATCCAGGATCAATCGAAAGATTC  792

Qy  230 PheAspGlnGluIleLeuValMetAlaLysCysValHisGluAsnLeuValGluLeuLeu  249
           |||:::               ::::::
Db  793 TTCCAGGCAGAGTTGCAGATTTGTCTTAGATGCTGCCACCCAAGTGTTCTTACCTGTGCTG  852

Qy  250 GlyPheSerSerAspGlyAspProLeuCysLeuValValTyrValTyrMetProAsnGlySer  269
           |||||||
Db  853 GGCTTCTGTCGTCCTGCAAGACAGTTCACAGCTTCATCTACCCCATACATGGCAAATGGTTCC  912

Qy  270 LeuLeuAspArgLeuSerCysLeuAspGlyThrProProLeuSerTrpHisMetArgCys  289
```

```
Db  913 CTACAGGACAGACTGCAGGGTCAGGGTGGCTCGGACCTTCCCGGCCCCAGCCGTGTC  972

Qy  290 LysIleLeuGlnGlyAlaAlaAlaAsnGlyIleAsnPheLeuHisGluAsnHisHisIleHis  309
           |||                :::                  :::||
Db  973 AGCATCTTGCTCAGGGCGGCTGCTCTCTGTGCCGTCCGAGTACCTGCATGGGTCGGGAGATCATCCAC  1032

Qy  310 ArgAspIleLysSerAlaAsnIleLeuLeuAspAspAlaPheThrAlaLysIleSerAsp  329
           ::::::|||||:::               ::::::|||
Db 1033 AGCAACGTCAAGAGCTTCTAAATGTCTTGGACCAAAAATCTCACCCCACCAAAACTTGCTCAC  1092

Qy  330 --------PheGlyLeuAlaArgGluGluLysProPheGluLeuLeu  347
                   :::               :::||
Db 1093 CCAATGGCTCATCTGTGTCCTTGTCAACAAAAGGTCAAAAATACACCATGATAAGACTCAC  1152

Qy  348 IleValGlyThrThr-----AlaTyrMetAlaProGluValAlaLeuArg---GlyGluIleThr  365
           :::|||                   :::               :::||
Db 1153 CTGCTCCGGACGGTCAGCCGCGGCGTTACCTGACGACTTTACTCCTCAGTGAAATTCCA  1212

Qy  366 ProLysSerAspIleLeuTyrSerPheGlyValValLeuLeuGluIleIleThrGlyLeuPro  385
           :::               |||:::||||::::::||||||
Db 1213 AAGCCAGATGGACATCTTTCAGCTGTGTGAGAATAGTGTTGGCCGGAGGTCCTCACGGGCATCCCT  1272

Qy  386 AlaValAlaAspGluHisArgGluProGlnLeuLeuLeuLeuAspIle----------  399
           :::||                :::               -----
Db 1273 GCAATGGATAACAACCGAAGCCCGGTTACCTGAAGGACTTACTCCTCAGTGAAATTCCA  1332

Qy  400 ------------------LysGluGluIleGluAsp------GluGluLys  408
                             |||               :::
Db 1333 AGCACACCGCCTCGGCTCTGCCTCCAGGAGAACCGCGGTGGAAGACGTGATAGGCAAAGGAG  1392

Qy  409 ThrIleGluAspTyrIleAspGlyLysLysMetAsnAlaAspSerThrSerValGluAla  428
           |||:::||||:::||              ||||||
Db 1393 ATCTGCCAGAAGTACCTGGAGAAGGGCCAGAGGAGGCTTCGGAGGACTGGCGCGCAGGCC  1452

Qy  429 MetTyrSerValAlaSerGlnCysLeuHisGluLysLysProAspIleLys  448
           |||               :::||
Db 1453 CTGGCCACGGCGTTGCCTGCCTGCGGAGGCGTAACCACCAGC-  1497

Qy  449 LysValGlnGluGlnLeuLeuGluMetThrAlaSer  460
           |||||||              |||
Db 1498 ---------CTGCAGGAGGTGTGTGGCTCT  1518
```

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 04:27:15 ; Search time 688.136 seconds
(without alignments)
2356.207 Million cell updates/sec

Title:
Perfect score: 2386
Sequence: 1 MNKPITPSTYVRCLNVGLIR..........KNKRPDIKKVQQLLQEMTAS 460

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152426_19815/app_query.fasta_1.1109
-DB=Published Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10001254 @CGN_1_1.534 @runat_16012004_152426_19815
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09C_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

1	2386	100.0	1383	15	US-10-001-254-15	Sequence 15, Appl
2	2382	99.8	1383	15	US-09-759-595-15	Sequence 2, Appli
3	2362	99.0	2817	11	US-09-966-451-3	Sequence 3, Appli
4	2362	99.0	2817	11	US-10-001-254-27	Sequence 27, Appl
5	2017	84.5	1542	10	US-10-001-254-4	Sequence 4, Appli
6	972	40.7	1493	10	US-09-764-868-249	Sequence 249, App
7	779	32.6	501	9	US-09-745-763-109	Sequence 109, App
8	691	29.0	501	9	US-09-833-790-149	Sequence 149, App
9	590	24.7	31000	11	US-09-966-451-10	Sequence 10, Appl
10	517	21.7	294	15	US-10-001-254-21	Sequence 21, Appl
11	515.5	21.6	3590	12	US-09-880-107-2222	Sequence 2222, Ap
12	515.5	21.6	3590	10	US-10-167-034-4	Sequence 4, Appli
13	515.5	21.6	3590	13	US-10-101-510-88	Sequence 88, Appl
14	515.5	21.6	3590	10	US-09-873-367C-1003	Sequence 1003, Ap
15	508.5	21.3	2361	10	US-09-938-842A-1673	Sequence 1673, Ap
16	506	21.2	1899	10	US-09-938-842A-799	Sequence 799, App
17	499	20.9	1281	10	US-09-938-842A-624	Sequence 624, App
18	498.5	20.9	2607	10	US-09-938-842A-1269	Sequence 1269, Ap
19	494	20.7	1389	10	US-09-938-842A-340	Sequence 340, App
20	485	20.3	1239	10	US-09-938-842A-1190	Sequence 1190, Ap
21	481	20.2	2196	14	US-10-086-464-16	Sequence 16, Appl
22	481	20.2	2261	14	US-10-086-464-15	Sequence 15, Appl
23	480.5	20.1	2640	10	US-09-938-842A-1270	Sequence 1270, Ap
24	479	20.1	1275	10	US-09-938-842A-1977	Sequence 1977, Ap
25	478.5	20.1	1944	14	US-10-086-464-1	Sequence 1, Appli
26	478.5	20.1	2189	14	US-10-086-464-3	Sequence 3, Appli
27	473.5	19.8	2733	15	US-10-101-464A-869	Sequence 869, App
28	472.5	19.8	3257	15	US-10-101-464A-455	Sequence 455, App
29	472.5	19.8	3257	15	US-10-101-464A-456	Sequence 456, App
30	471	19.7	1257	13	US-10-342-224-67	Sequence 67, Appl
31	470	19.7	3453	13	US-10-259-165-311	Sequence 311, App
32	468.5	19.6	2690	15	US-10-101-464A-870	Sequence 870, App
33	468.5	19.6	2466	10	US-09-938-842A-2644	Sequence 2644, Ap
34	466.5	19.6	1098	10	US-09-938-842A-2233	Sequence 2233, Ap
35	459.5	19.3	3237	15	US-10-101-464A-858	Sequence 858, App
36	457.5	19.2	1419	10	US-09-938-842A-926	Sequence 926, App
37	456	19.1	2124	10	US-09-938-842A-1079	Sequence 1079, Ap
38	456	19.1	2144	13	US-10-304-946-7	Sequence 7, Appli
39	456	19.1	2242	13	US-10-304-946-6	Sequence 6, Appli
40	456	19.1	2411	13	US-10-304-946-8	Sequence 8, Appli
41	456	19.1	2736	10	US-09-938-842A-56	Sequence 56, Appl
42	456	19.1	70768	14	US-10-135-322-13	Sequence 13, Appl
43	455	19.1	2685	11	US-09-754-853A-7	Sequence 7, Appli
44	455	19.1	2685	11	US-09-754-853A-50	Sequence 50, Appli
45	455	19.1	2685	11	US-09-754-853A-51	Sequence 51, Appli

ALIGNMENTS

RESULT 1
US-10-001-254-15

Sequence 15, Application US/10001254
Publication No. US20030049702A1
GENERAL INFORMATION:
APPLICANT: Reed, John C.
APPLICANT: Godzik, Adam
APPLICANT: Pawlowski, Krzysztof
APPLICANT: Fiorentino, Loredana
APPLICANT: Lee, Sug Hyung
APPLICANT: Roth, Wilfred
APPLICANT: Stenner-Liewen, Frank
TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins
FILE REFERENCE: P-LJ 5037
CURRENT APPLICATION NUMBER: US/10/001,254
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: 60/301,889
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: 09/715,893
PRIOR FILING DATE: 2000-11-17
NUMBER OF SEQ ID NOS: 62
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 15
LENGTH: 1383

```
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1380)
US-10-001-254-15

DB:
Query Match:             15
Best Local Similarity:  100.00%
Percent Similarity:     100.00%
Percent Identity:       100.00%
Score:                  2386.00

Alignment Scores:
Pred. No.:

US-10-001-254-16 (1-460) x US-10-001-254-15 (1-1383)

Score:    3.75e-259        Length:      1383
                           Matches:      460
                           Conservative:   0
                           Mismatches:     0
                           Indels:         0
                           Gaps:           0
```

```
Qy     1  MetAsnLysProIleThrProSerThrThrTyrValArgCysLysLeuAsnValGlyLeuIleArg   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ATGAACAAACCCATAACACCATCAACAACATATGTGCGGTGCTGCTCAAGTTGAACGTTGGA     60

Qy    21  LysLeuSerAspAspPheIleAspProGlnGlnGlyTrpLysGlyLysGluAlaValAlaIleLeuLys   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAAGGAAAGTTAGCTGTTAGCTAAA    120

Qy    41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgArgThrIleGluAlaLeuLeu    60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  AAACCATCTGGTGATGATAGATACAATCAGTTTCATAAAGAGGAGTTGAAGCATTACTT    180

Qy    61  GlnThrGlyLysSerProThrSerGluLeuLeuPheGluAspLeuGlyTyrThrThrAsnCysThr    80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CAAACTGGAAAAAGTCCCACCTCTGAATTACTGTTTGAACTGGGCCACCACAAATTGCACA   240

Qy    81  ValGlyAspLeuValAspGlnLeuLeuIleLeuAsnGluPhePheAlaProAlaSerLeuLeu   100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GTTGGTGATCTTGTGGGATCTTTTGATCCAAAATGAATTTTTTGCTCCGCGAGTCTTTTG    300

Qy   101  LeuProAspAlaValProLysSerThrAlaAsnThrLeuProSerLysGluAlaIleThrVal   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  CTCCCAGATGCTGTTCCCAAAACTGCTAATACACATACACCTTCTAAAGAAGCTATAACAGTT   360

Qy   121  GlnGlnLysGlnMetProPheCysSerAspLysAspSerArgThrLeuMetThrProValGlnAsn   140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  CAGCAAAAACAGATGCCTTTCTGTCAAGACAAAGACAGCAGACATTGATGACACCTGTGCAGAAT   420

Qy   141  LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerSerLeuGluVal   160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  CTTGAACAAAGCTATATGCCACCTGACTCCTCAAGTCCAGAAAATAAAAGTTTAGAAGTT    480

Qy   161  SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGAAGAATGTCACAAATAACTTT   540

Qy   181  AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVal   200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  GATGAACGACCAATTTCTGTTGGTGGTAATAAAATGGGAGAGGGAGGATTTGGAGTTGTA   600

Qy   201  TyrLysGlyTyrValAsnAsnThrThrValAlaValLysLysLeuAlaAlaMetValAsp   220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  TATAAAGGCTACGTAAATAACACAACTGTGGCAGTGAAGAAGCTTGCAGCAATGGTTGAC   660

Qy   221  IleThrThrGluGluLeuLysGlnGlnPheAspGlnGluIleLysValMetAlaLysCys   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  ATTACTACTGAAGAACTGAAACAGCAGTTTGATCAAGAAATAAAAGTAATGGCAAAGTGT   720

Qy   241  GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeuCysLeu   260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  CAACATGAAAACTTAGTAGAACTACTGGGTTTCTCAAGTGATGGAGATGACCTCTGCTTA   780

Qy   261  ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr   280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  GTATATGTTTACATGCCTAATGGTTCATTGCTTGACAGACTCTCTTGCTTGGATGGTACT   840
```

```
RESULT 2
US-09-759-595-2

; Sequence 2, Application US/09759595
; Publication No. US20030059915A1
; GENERAL INFORMATION:
;   APPLICANT: Wesche, Holger
;   APPLICANT: Li, Shyun
;   APPLICANT: Tularik Inc.
;   TITLE OF INVENTION: IRAK-4: Compositions and Methods of Use
;   FILE REFERENCE: 018781-00391OUS
;   CURRENT APPLICATION NUMBER: US/09/759,595
;   CURRENT FILING DATE: 2001-01-13
;   PRIOR APPLICATION NUMBER: US 60/176,395
;   PRIOR FILING DATE: 2000-01-13
;   NUMBER OF SEQ ID NOS: 7
;   SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1383
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human IL-1 receptor-associated kinase 4 (IRAK-4)
;   FEATURE:
;   OTHER INFORMATION: cDNA
;   NAME/KEY: CDS
;   LOCATION: (1)..(1383)
;   OTHER INFORMATION: human IRAK-4
US-09-759-595-2

DB:
Query Match:             11
Best Local Similarity:   99.83%
Percent Similarity:      99.78%
Percent Identity:        99.78%
Score:                   2382.00

Alignment Scores:
Pred. No.:

Score:    1.06e-258        Length:      1383
                           Matches:      459
                           Conservative:   0
                           Mismatches:     0
                           Indels:         1
                           Gaps:           0
```

```
Qy   281  ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  CCACCACTTTCTTGGCACATGAGATGCAAGATCCAAGATTGCTACAGGGTGCAGCTAATGGCAT    900

Qy   301  PheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp   320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  TTTCTACATGAAAAATCATCATATTCATAGAGATATTAAAAGTGCAAAATCTTACTGGAT    960

Qy   321  GluLeuAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla   340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  GAAGCTTTTACTGCTAAAATATCTGACTTTGGCCTTGCAAGAGCAAGTGAAAAGTTTGCC   1020

Qy   341  GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  CAGACAGTCATGACTAGCAGAATTGTGGGAACAACAGCTTATATGGCACCAGAAGCTTTG   1080

Qy   361  ArgGlyGlyIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle   380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  CGTGGAGGAAATAACACCCAAAGTGATATTTACAGCTTTGGTGTGGTTTTACTAGAAATA   1140

Qy   381  IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAspIleLys   400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  ATAACTGGACTTCCAGCTGTGGATGAACACCGTGAACCTCAGTTATTGCTAGATATTAAA   1200

Qy   401  GluGluLeuGluAspGluGluLysThrIleGluAspTyrIleAspLysLysMetAsnAsp   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  GAAGAATTGGAAGATGAAGAAAAGACAATTGAAGATTATATTGATAAAAAGATGAATGAT   1260

Qy   421  AlaAspSerThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys   440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  GCTGATTCCACTTCAGTTGAAGCTATGTACTCTGTTGCTAGTCAGTGCTTGCATGAAAAG   1320

Qy   441  LysAsnLysArgProAspIleLysLysValGlnGlnLeuLeuGlnGluMetThrAlaSer   460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  AAAAATAAGAGACCAGACATTAAGAAAGTTCAACAGCTGCTGCAAGAGATGACAGCTTCT   1380
```

US-10-001-254-16 (1-460) x US-09-759-595-2 (1-1383)

```
Qy      1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
Db      1 ATGAACAAACCCATAACACCATCAACACATATGTGCGCTGCCTCCAAATGTTGGACTAATTAGG 60

Qy     21 LysLeuSerAspPheIleAspProGlnGlyTrpLysLysLeuAlaValAlaAlaIleLys 40
Db     61 AAGCTGTCAGATTTTATTGATCCTCAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 120

Qy     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
Db    121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGATTTGAAGCATTACTT 180

Qy     61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
Db    181 CAAACTGGAAAAAAGTCCCCACTTCTGAATTACTGTTTGACTGGGGCACCAACAATTACTT 240

Qy     81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPheAlaLeuAlaProAlaSerLeuLeu 100
Db    241 GCTGGTGATCTTGTGGATCTTTTGATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTG 300

Qy    101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120
Db    301 CTCCCAGATGCTGTTCCCAAAAACTGCTAATACACTTCCTTCTAAAGAAGCTATAACAGTT 360

Qy    121 GlnGlyLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn 140
Db    361 CAGGGAAAAACAGATGCCTTCTGTGACAAAGACAGACATTGACACCTGTCAGAAT 420

Qy    141 LeuGlyLysGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal 160
Db    421 CTTGGACAAAAGCTATATGCCACCTGACTCCTCAAGTCCAGAAAATAAAAGTTTAGAAGTT 480

Qy    161 SerAspThrArgPheHisSerPheSerPheThrTyrGluLeuLeuLysAsnValThrAsnAsnPhe 180
Db    481 AGTGATACACGTTTTCACAGTTTTCATTTTATGAATTGAAATGCACAAATAACTTT 540

Qy    181 AspGluArgProIleSerGluValGlyIleGlyValGlyValPheValVal 200
Db    541 GATGAACGACCCATTTCTGTTGGTGTAATAAAATGGGAGGAGGATTTGGAGTTGTA 600

Qy    201 TyrLysGlyTyrValAsnSerThrThrValAlaValAlaValLysLysLeuAlaAlaMetValAsp 220
Db    601 TATAAAGGCTACGTAAATAACACAACTGTGGCAGTGAAGAAGCTTGCAGCAATGGTTGAC 660

Qy    221 IleThrThrGluGluLeuLysGlnPheAspProGlnLeuValMetAlaLysCys 240
Db    661 ATTACTACTGAAGAACTGAAACAGTTTGATCCAAGAAAATAAGTAATGGCAAAGTGT 720

Qy    241 GlnHisGluAsnLeuValGluLeuGlyPheSerSerAspGlyAspLeuCysLeu 260
Db    721 CAACATGAAAAACTTAGTAGAACTACTGGTTTCTCAAGTGATGGAGATACCTCTGCTTA 780

Qy    261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspSerGlyThr 280
Db    781 GTATATGTTTACATGCCTAATGGTTCATTGCTAGACAGACTCTCTTGCTTGGATAGTGGTACT 840

Qy    281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyValAlaAlaAsnGlyIleAsn 300
Db    841 CCACCACTTCTTGGCACATGAGATGCCAAGATTGCTCAAGGGTGCAGCTAAGGCATCAAT 900

Qy    301 PheLeuHisGluAsnLeuValGlyPheSerAlaAsnIleLeuLeuAsp 320
Db    901 TTTCTACATGAAAATCTATCATATTCATGAGAATTTACTGGAT 960

Qy    321 GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla 340
Db    961 GAAGCTTTACTGCTAAAAATCTGACTTTGGCCTTGACACGGGCTTCTGAGAAGTTTGCC 1020

Qy    341 GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu 360
Db   1021 CAGACAGTCATGACTAGCAGAATTGTGGGGACAACAGCTTATATGGCACCAGAAGCTTTG 1080
```

RESULT 3
US-09-966-451-3
```
; Sequence 3, Application US/09966451
; Publication No. US20030087856A1
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF IL-1 RECEPTOR-ASSOCIATED KINASE-4 EXPRES
; FILE REFERENCE: RTS-0324
; CURRENT APPLICATION NUMBER: US/09/966,451
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 3
; LENGTH: 2817
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (50)...(1432)
US-09-966-451-3
```

Alignment Scores:
```
Pred. No.:
Score:                    5.62e-256    Length:              2817
Score:                    2362.00      Matches:             456
Percent Similarity:       99.13%       Conservative:        0
Best Local Similarity:    99.13%       Mismatches:          4
Query Match:              98.99%       Indels:              0
DB:                       11           Gaps:                0
```

US-10-001-254-16 (1-460) x US-09-966-451-3 (1-2817)

```
Qy      1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
Db     50 ATGAACAAACCCATAACACCATCAACACATATGTGCGCTGCCTCCAAATGTTGGACTAATTAGG 109

Qy     21 LysLeuSerAspPheIleAspProGlnGlyTrpLysLysLeuAlaValAlaAlaIleLys 40
Db    110 AAGCTGTCAGATTTTATTGATCCTCAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 169

Qy     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
Db    170 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGATTTGAAGCATTACTT 229

Qy     61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
Db    230 CAAACTGGAAAAAAGTCCCCACTTCTGAATTACTGTTTGACTGGGGCACCAACAATTACACA 289

Qy     81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPheAlaLeuAlaProAlaSerLeuLeu 100
Db    290 GTTGGTGATCTTGTGGATCTTTTGATCCAAAATGAATTTTTGCTCCTGCGAGTCTTTTG 349

Qy    101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120
Db    101 ... LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120
```

RESULT 4
US-10-001-254-27

Sequence 27, Application US/10001254
Publication No. US20030049702A1
GENERAL INFORMATION:
  APPLICANT: Reed, John C.
  APPLICANT: Godzik, Adam
  APPLICANT: Pawlowski, Krzysztof
  APPLICANT: Fiorentino, Loredana
  APPLICANT: Lee, Sug Hyung
  APPLICANT: Roth, Wilfried
  APPLICANT: Stenner-Liewen, Frank
  TITLE OF INVENTION: No. US20030049702A1el Death Domain Proteins
  FILE REFERENCE: P-LJ 5037
  CURRENT APPLICATION NUMBER: US/10/001,254
  CURRENT FILING DATE: 2001-11-15
  PRIOR APPLICATION NUMBER: 60/301,889
  PRIOR FILING DATE: 2001-06-29
  PRIOR APPLICATION NUMBER: 09/715,893
  PRIOR FILING DATE: 2000-11-17
  NUMBER OF SEQ ID NOS: 62
  SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 27
  LENGTH: 2817
  TYPE: DNA
  ORGANISM: Homo sapien
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (50)...(1429)
US-10-001-254-27

US-10-001-254-16 (1-460) x US-10-001-254-27 (1-2817)

Alignment Scores:
Pred. No.:
Score: 5.62e-256   Length: 2817
Percent Similarity: 2362.00   Matches: 456
Best Local Similarity: 99.13%   Conservative: 0
Query Match: 99.13%   Mismatches: 4
DB: 98.99%   Indels: 0
                        15   Gaps: 0

Qy   1   MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1   ATGAACAAACCCATAACACCATCAACACATATGTGCGCTGCCTCAACGTGGGACTAATTAGG   109

Qy   21   LysLeuSerAspPheIleAspProGlnGlyTrpLysLysLeuAlaValAlaIleIleLys   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50   ATGAACAAACCCATAACACCATCAACACATATGTGCGCTGCCTCAACGTGGGACTAATTAGG   109

Qy   41   LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110   AAGCTGTCAGATTTTATTGATCCCCAAGGAAGGGATGGAAGAAGTTAGCTGTATTAAA   169

Qy   61   GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 170   AAACCATCTCGGTGATGATAGATACAATCACGTTTCACATAAGGAGATTTGAAGCATTACTT   229

Qy   81   ValGlyAspLeuValAspLeuLeuIleGlnAsnGluAlaAlaSerLeuLeu   100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 230   CAAACTGGAAAAAAGTCCCACTTCTGAATTACACGTTTTGACTGGGGGCACCACAAATTGCACA   289

Qy 101   LeuProAspAlaValProLysThrAlaAsnThrLeuProSerGlyLysAlaIleThrVal   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 290   GTTGGTGATCTTGTGGATCTTTTGATCCAAAATGAAATTTTTTGCTCCCTGCGAGTCTTTTG   349

Qy 121   GlnGlnLysGlnMetProPheCysAlaAspLysAspArgThrLeuMetThrProValGlnAsn   140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 350   CTCCCAGATGCTGTTCCCAAAAACTGCTAATACACTAACCTTCTAAAGGAAGCTATAACAGTT   409

Qy 141   LeuGluGlnSerTyrTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal   160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 410   CAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGGACATTGACACCTGTGCAGAAT   469

Qy 161   SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 470   CTTGAACAAAGCTATATGCCACCTGACTCCTCAGAAAATAAAAGTTTAGAAGTT   529

US-10-001-254-16 (1-460) x US-09-759-595-4 (1-1542)

Alignment Scores:

Pred. No.:	2.02e-217	Length:	1542
Score:	2017.00	Matches:	385
Percent Similarity:	90.85%	Conservative:	32
Percent Identity:	83.08%	Mismatches:	42
Best Local Similarity:	84.53%	Indels:	0
Query Match:	11	Gaps:	0
DB:			

```
Qy      1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg   20
           |||||||||||||||||||||||||||::|||||||||||||||||||||||||||||||
Db    163  ATGAACAAGCCCGTTCGACACCATCGACATACAATACGCAACCTTAATGTGGGGATCCTTAGG  222

Qy     21  LysLeuSerAspPheIleAspProGlnGluGlyTrpLysValGlyValAlaValAlaIleLys   40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db    223  AAGCTGTCCGGATTTTATTGATCCTCAAGAAGGGTGGAAGAAATTAGGAGTAGCTGTAGCTATCAAA  282

Qy     41  LysProSerGlyAspAspArgTyrAsnGlnPheHisGlyIleAlaArgPheGluAlaLeuLeu   60
           |||||||||||||||||||||||||||||::||||||||||||||||||||||||||
Db    283  AAGCCCGGTGGCGACGACAGATACAATCAGTTCCATATAAGGAGCCTTACTT  342

Qy     61  GlnThrGlyLysSerProThrSerGluLeuLeuPheGlyThrThrAsnCysThr   80
           |||||||||||||||||||||||||||||||||||||::|||||||||||||||||
Db    343  CAGACCGGGAAGAGCCCCACCTGTGAACTGCTGTTTGGCACTACGAACTGCACA  402

Qy     81  ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeu  100
           |||||||||||||||||||||||||||||:||||||||||||||||:::||||||||||
Db    403  GTTGGGGACCTTGTGGATCTACTGGTCCAGAATGAGCTGTTTGCCCCGCCACTCTCCTG  462

Qy    101  LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysValAlaIleThrVal  120
           ||||||||||||||||||||||::::|||||||||::||||||||||||||||
Db    463  CTGCCCGATGCCGTTCCCAAAACCGTCAAAACGCTCCCTCTAGAAGAAGGCCGCAACAGTG  522

Qy    121  GlnGlnLysGlnMetProPheCysSerAspPheLysAspArgThrLeuMetThrProValGlnAsn  140
           ||||||||||||||||||||||||||||||||||||:::||
Db    523  GCACAAAACACACGGGGCCTTGTCAGGAAAAGGACAGCATCCTTATGCCCAAG  582

Qy    141  LeuGluGlnSerTyrMetProProAspSerSerSerProGluGlnAsnLysSerSerLeuGluVal  160
           ||||||||||||||||||||||||||||||:::||||||||||||||::||
Db    583  CTAGAACAGAGCTACGCAGCCGAGCCATCCTCCAGCACAACAGAAGTGTAGAGTCC  642

Qy    161  SerAspThrArgPheHisSerPhePheTyrGluLeuLeuLysAsnValThrAsnAsnPhe  180
           ||||||||||||||||||||||||||||:::::::|||||
Db    643  AGCGACACCTCGGTTCCATGAAGAACTGAAGAGCATCACAAAACAACTTC  702

Qy    181  AspGluArgProIleSerValGlyAsnLysMetGlyGluGlyPheGlyValVal  200
           |||||||||||||||||||||||||||||||||||||||||
Db    703  GACGAGCGAACCCGCGTCTGCCGGTGCAACCGGATGGGAGTGGTG  762

Qy    201  TyrLysGlyTyrValAsnAsnThrThrValAlaValAlaLysLysLeuAlaAlaMetValAsp  220
           ||||||||||||||||||||||||||||||||||||||
Db    763  TACAAGGGCTGTGTGAACAACACCATACGTGCGGTGGCGGGAGAGCTCCGGACGGCGGATTGAA  822

Qy    221  IleThrThrGluGlnLeuLysGlnHisPheAspGlnGluIleLysValMetAlaLysCys  240
           ||||::||||||:|||||||||||||||||||||||||||:|
Db    823  ATCAGTACTGAAGAACTAAAAGCAACAGTTTGATCAAGAAATTAAAGTGGCAAGTGT  882

Qy    241  GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspLeuCysLeu  260
```

```
Db    530  AGTGATACACGTTTTCACAGTTTTCATTTATGAATTGAGAAATGTCACAAATACTTT  589

Qy    181  AspGluArgProIleSerValGlyAsnLysMetGlyGluGlyPheGlyValVal  200
           |||||||||||||||||||||||||||||||||||
Db    590  GATGAGCGACCCCATTTCTGTTGGTGGTAATAAAATGGGAGAGGGAGGATTTGGAGTTGTA  649

Qy    201  TyrLysGlyTyrValAsnAsnThrThrValAlaValAlaLysLysLeuAlaAlaMetValAsp  220
           |||||||||||||||||||||||||||||||||
Db    650  TATAAAGGCTACGTAAATAACACAACTGTGGCAGTGAGAAGCTTGAAGCTTGCAGCAATGGTTGAC  709

Qy    221  IleThrThrGluGlnLeuLysGlnHisPheAspGlnGluIleLysValMetAlaLysCys  240
           |||||||||||||||||||||||||||||
Db    710  ATTACTACTGAAGAACTGAAACAGCACAGTTTGATCAAGAAAATAAAAGTAATGCAAAGTGT  769

Qy    241  GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspLeuCysLeu  260
           |||||||||||||||||||||||||||||||
Db    770  CAACATGAAAACTTAGTAGAACTACTTGGTTTCTCAAGTGATGGAGACCTCTGCTTA  829

Qy    261  ValThrTyrValThrTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr  280
           ||||||||||||||||||||||||||||||||||||
Db    830  GTTACATATGTTACATGCCTAATGGTTCATTGCTACAGACAGCTCTTCGCTTGGGATGGTACT  889

Qy    281  ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn  300
           ||||||||||||||||||||||||||||||||
Db    890  CCACCACTTCTTGGCACATGAGATCAAGATTGCTCAGGGTGCACCTAATGGCCATCAAT  949

Qy    301  PheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAsp  320
           |||||||||||||||||||||||||||||||
Db    950  TTTCTACATGAAAATCATCATATTCATAGAGATATTAAAAGTGCAAATATCTTACTGGAT  1009

Qy    321  GluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla  340
           ||||||||||||||||||||||||||||||
Db   1010  GAAGCTTTTACTGCTAAAATATCTGACTTTGGCCTTGCACGGGCTTCTGAGAAGTTTGCC  1069

Qy    341  GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu  360
           ||||||||||||||||||||||||||||||
Db   1070  CAGACAGTCATGACTAGCAGAATTGTGGGAACAACAGCTTATATGGCACCAGAAGCTTTG  1129

Qy    361  ArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle  380
           ||||||||||||||||||||||||||||||
Db   1130  CGTGGAGAAATAACACCCAAATCTGATATTTACAGCTTTGGTGTTGTTTTACTAGAAATA  1189

Qy    381  IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAspIleLys  400
           ||||||||||||||||||||||||||||||
Db   1190  ATAACTGGACTTCCAGCTGTGGATGAACACCGTGAGCCTCAGTTATTGCTAGATATTAAA  1249

Qy    401  GluGluIleGluAspGluGluLysThrIleGluAspTyrIleAspLysLysMetAsnAsp  420
           |||||||||||||||||||||||||||||:::|
Db   1250  GAAGAAATTGAAGATGAAGAAAAGACAATTGAAGATTATATTGATAAAAAGATGAATGAT  1309

Qy    421  AlaAspSerThrSerValGlyAlaAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys  440
           ||||||||||||||||||||||||||||||
Db   1310  GCTGATTCCACTTCAGTTGGAGCTAGTGTACTTCGGGTACTCGGCCAATGTCGGCATGAAAAG  1369

Qy    441  LysAsnLysSerArgProAspIleLysLysValGlnLeuLeuGlnLeuGluMetThrAlaSer  460
           ||||||||||||||||||||||||||||||
Db   1370  AAAAATAAGAGCCCAGACATTAAGAAGGTTCACCAGCTGCTGCAAGATGACAGCTTCT  1429
```

```
RESULT 6
US-09-764-868-249/c
Sequence 249, Application US/09764868
Patent No. US20020168711A1
GENERAL INFORMATION:
    APPLICANT: Rosen et al.
    TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
    FILE REFERENCE: PTZ32
    CURRENT APPLICATION NUMBER: US/09/764,868
    CURRENT FILING DATE: 2001-01-17
    Prior application data removed - refer to PALM or file wrapper
    NUMBER OF SEQ ID NOS: 1510
    SOFTWARE: Patentin Ver. 2.0
    SEQ ID NO 249
    LENGTH: 1493
    TYPE: DNA
    ORGANISM: Homo sapiens
US-09-764-868-249
```

```
Alignment Scores:
Pred. No.:
Score:                    1.866e-99        Length:        1493
Percent Similarity:       100.00%          Matches:        190
Percent Identity:         100.00%          Conservative:     0
Best Local Similarity:     40.74%          Mismatches:       0
Query Match:                  10           Indels:           0
DB:                                        Gaps:             0
```

US-10-001-254-16 (1-460) x US-09-764-868-249 (1-1493)

```
Qy   271  LeuAspArgLeuSerCysLeuAspGlyThrProProLeuSerTrpHisMetArgCysLys  290

Db   883  CAGCACGAGGAGAACCTGGGTGGAGCTGCTCGGGCTTCTCCAGCGACGAGGACAAACCTGTGCTTA  942

Qy   261  ValTyrValTyrMetProAsnGlySerLeuLeuSerCysLeuAspGlyThr            280
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   943  GTGTATGCTTACATGCCAAACGGGTCCTTGCTGAGCTGCCTGGACGGTACA           1002

Qy   281  ProProLeuSerTrpHisMetArgCysLysSerIleAlaGlnGlyAlaAlaAsnGlyIleAsn  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1003  CCACCGCGCTTCCTGCGACACAGAAGTGCAAGAGTGCAGCAGCAAAATGGGCATCAGG     1062

Qy   301  PheLeuHisGluAsnHisHisIleHisArgAspGlySerAlaAsnIleLeuLeuAsp       320
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1063  TTTCTGCATGAAAATCATCACATTCATAGAGATGGGTCAGCAAATATTCTTACTAGAC     1122

Qy   321  GlyLeuAlaPheThrAlaLeuSerAspPheGlyLeuLeuAlaArgGluGlyLysPheAla     340
          |||||||||||||||||||||||||||||||||||||||||||||||||||  :::
Db  1123  AAAGGACTTTACTGCCAAAATATTCTGACTTTGGGCTTGCACGGGAAGGCTAGCG        1182

Qy   341  GlnThrArgValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1183  CAGACAGGTCATGACCAGCCAACGGGCCTTACATGGCACACCGAAGCCTTTG          1242

Qy   361  ArgGlyGlyIleThrProLysSerAspAspIleTyrSerPheGlyValValLeuLeuGluIle 380
          |||||||||||||||||||||||||||||||||||||||||||::::::
Db  1243  CGGGGAGGAAATAACACCCAAAATCTGACATTACACTTCCGGCTGTGGTTCTTGGAGCTG   1302

Qy   381  IleThrGlyLeuProAlaValAlaAspPheHisArgGluProGlnLeuLeuLeuAspIleLys 400
          ||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db  1303  ATAACCGGGCTGCCGGCTGTGGATGAAAACCGTGAACCTCAACTACTGCTGGATATTAAA   1362

Qy   401  GluGluIleGluAspGluGluLysTyrThrIleGluAspTyrIleLeuLysSerMetAsnAsp 420
          |||||||||||||||||||||||||||||||||||||:::||||||
Db  1363  GAAGAGATTGAAGATGAAGAGAAGTACACGATTGAAGACTATATTCTGAAAAGTATGAATGAT 1422

Qy   421  AlaAlaAspSerThrSerValGlyGluAlaMetTyrSerValAlaSerGlnCysLeuHisGly 440
          |||||||||||||||||||||||||||||||||||||||::::::
Db  1423  GCGGCCGACTCCACCTCGGTGGGAGAAGCAATGTACTCCGTGGCTTCCCAGTGTCTGCATGGAG 1482

Qy   441  LysAsnLysArgProAspIleLeuLysSerValGlnIleLeuGlnGluMetThrAla        459
          ||||||||||||||||||||||||:::|||
Db  1483  AAAAACAAGAGGCCAGACATTGCAAAGGTTCAACAGCTGCTACAAGAGATGTCTGCT        1539
```

```
RESULT 7
US-09-745-763-109
Sequence 109, Application US/09745763
Patent No. US20020065394A1
GENERAL INFORMATION:
    APPLICANT: Jacobs, Kenneth
               McCoy, John M.
               LaVallie, Edward R.
               Collins-Racie, Lisa A.
               Evans, Cheryl
               Merberg, David
               Treacy, Maurice
               Spaulding, Vikki
    TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
                        ENCODING THEM
    NUMBER OF SEQUENCES: 219
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Genetics Institute, Inc.
        STREET: 87 CambridgePark Drive
        CITY: Cambridge
        STATE: MA
        COUNTRY: U.S.A.
        ZIP: 02140
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/745,763
        FILING DATE: 18-Jun-2000
        CLASSIFICATION: <Unknown>
    ATTORNEY/AGENT INFORMATION:
        NAME: Springer, Suzanne A.
```

```
Qy   271  LeuAspArgLeuSerCysLeuAspGlyThrProProLeuSerTrpHisMetArgCysLys  290

Db  1491  CTAGACAGACTCTCTTGCTTGGAGTGGTACTCCACCACTTTCTTGGCACATGAGATGCAAG 1432

Qy   291  IleAlaGlnGlyAlaAlaAlaGlyGlyGlyThrIleIleAsnPheLeuHisGluAsnHisHisIleHisArg 310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1431  ATTGCTCAGGGTGCAGCTGCAGGTGGAGGCATCATAAATTTCTACATGAAAAATCATCATATTCATAGA 1372

Qy   311  AspIleLysSerAlaAsnIleLeuLeuAspGlyLeuAlaArgAlaSerAspPhe           330
          |||||||||||||||||||||||||||||||||||:::
Db  1371  GATATTAAAAGTGCAAAATATCTTACTGGATGAAGGTTTTACTGCTCGTGCAAGTGACTTT 1312

Qy   331  GlyLeuAlaArgAlaSerGluGlyLysPheAlaGlnThrArgValMetThrSerArgIleValGly 350
          |||||||||||||||||||||||||||||||||||||||
Db  1311  GGCCTTGCACGGGCTTCTGAGGAAGTTTGCCCAGACAGTCATGACTAGCCAGAATTGTGGGA 1252

Qy   351  ThrThrAlaTyrMetAlaProGluAlaLeuArgGlyGlyIleThrProLysSerAspAspIle 370
          |||||||||||||||||||||||||||||||||||||||
Db  1251  ACAACAGCTTATATGGCACCAGAAGCCTTTCGGGGAGGAAATAACACCCAAAATCTGATATT 1192

Qy   371  TyrSerPheGlyValValLeuLeuGluIleIleThrGlyLeuProAlaValAlaAspHis    390
          |||||||||||||||||||||||||||||:::
Db  1191  TACAGCTTTGGTGTGGTTTTACTAGAAATAATAACTGGACTTCCAGCTGTGGCTGACCAC   1132

Qy   391  ArgGluProGlnLeuLeuLeuAspIleLysGluGluIleGluAspGluGluLysTyrThrIle 410
          ||||||||||||||||||||||||||||:::
Db  1131  CGTGAACCTCAGTTATTGCTAGATATTAAAGAAGAGATTGAAGATGAAGAAAAGTACACAATT 1072

Qy   411  GluAspTyrIleLeuLysSerMetAsnAspAlaAlaAspSerThrSerValGlyGluAlaMetTyr 430
          |||||||||||||||||||||||||||||||||||:::
Db  1071  GAAGATTATATTGATAAAAGATGAATGATGCTGATTCCACTTCAGTTGAAGCTATGTAC    1012

Qy   431  SerValAlaSerGlnCysLeuHisGlyLysAsnLysArgProAspIleLeuLysSerVal    450
          ||||||||||||||||||||||||||||||||||||
Db  1011  TCTGTTGCTAGTCAATGTCTGCATGAAAAGAAAAATAAGAGACCAGACATTAAGAAGGTT    952

Qy   451  GlnGlnIleLeuGlnGluMetThrAlaSer                                 460
          |||||||||||||||||||||
Db   951  CAACAGCTGCTGCAAGAAGATGACAGCTTCT                                922
```

; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 109:
; US-09-745-763-109

Alignment Scores:
Pred. No.:
Score: 779.00
Percent Similarity: 98.76%
Best Local Similarity: 98.76%
Query Match: 32.65%
DB:

```
                 Length:     483
                 Matches:    159
                 Conservative: 0
                 Mismatches:  1
                 Indels:      2
                 Gaps:        0
```

US-10-001-254-16 (1-460) x US-09-745-763-109 (1-483)

```
Qy  270 LeuLeuAspArgLeuSerCysLeuAspGlyThrProProLeuSerTrpHisMetArgCys 289
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3 TTGCTAGAGACAGACTCTCTTGCTTGGATGGTACTCCACCACTTTCTTGGCACGATGAGATGC 62

Qy  290 LysIleAlaGlnGlyAlaAlaAsnGlyIleIleAsnPheLeuHisGlnAsnHisHisIleHis 309
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 AAGATTGCTCAGGGTGCAGCTAATGGCATCAATTTTCTACATGAAAATCATCATATTCAT 122

Qy  310 ArgAspIleGlyLysSerAlaLaAsnIleLeuLeuAspGluAlaPheThrAlaLysIleSerArgAsp 329
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123 AGAGATATTAAAAGTGCAAAATACTTTACTGGAGGCTTTTACTGCTAAAAATATTCTGAC 182

Qy  330 PheGlyLeuLeuAlaArgAlaSerGluLys-PheAlaGlnThrValMetThrSerArgIleVa 349
        |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  183 TTTGGGCCTTGCACGGGCGTTCTGAGAAGTTTTGCCCAGACAGTTCATGACTAGCAGAATTGT 242

Qy  349 lGlyThrThrAlaTyrMetAlaProGlnAlaLeuAlaArgGlyGlyIleIleThrProLysSerAs 369
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  243 GGGACAACAACACCTTATATGGCACCAGAAGCTTTGCGTGGAGAAATAACACCCAAAATCTGA 302

Qy  369 pIleTyrSerPheGlyValValLeuLeuGluLeueIleThrGlyLeuProLeuAlaValAspGl 389
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  303 TATTTACAGCTTTGGTGTGGTTTTACTAGAAAATAACTGGACTTCCAGCTGTGGAATGA 362

Qy  389 uHisArgGluProGlnLeuLeuLeuAspIleLysGluLysGluGluGlyLysTh 409
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  363 ACACCCGTGAACCTCAGTTATTGCTAGATATTAAAGAAGAAAATGAGAGAGAAAGAC 422

Qy  409 rIleGluAspTyrIleAspLysMetAspSerThrSerValGluAlaMe 429
        |||||||||||||||||||||||||||||||||||||||||||| |||
Db  423 ATT-GAAGATTATATTGATAAAGATGAATGCTGATTCCACTTCAGTTGAAGCTAT 481

Qy  429 t 429
          |
Db  482 G 482
```

; FILE REFERENCE: 210121,512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(501)
; OTHER INFORMATION: n = A,T,C or G
; US-09-833-790-149

Alignment Scores:
Pred. No.:
Score: 691.00
Percent Similarity: 94.44%
Best Local Similarity: 93.21%
Query Match: 28.96%
DB:

```
                 Length:     501
                 Matches:    151
                 Conservative: 0
                 Mismatches:  2
                 Indels:      7
                 Gaps:        0
```

US-10-001-254-16 (1-460) x US-09-833-790-149 (1-501)

```
Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuLeuArg 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ATGAACAAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 80

Qy   21 LysLeuSerArgAspPheIleAspProGlnGluGlyTrpProLysGlyLeuAlaValAlaIleLys 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGCCAAAGGGATTAGGTGTAGCTATTAAA 140

Qy   41 LysProSerGlyAspAspArgTyrAsnGln-PheHisIleArgArgPheGluAlaLeuLeu 60
        |||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Db  141 AAACCATCTGGTGATGATAGATCACAATCAAGTTTCACATAAGGAGAGATTTGAAGCATT-CT 199

Qy   60 uGlnThrGlyTyrSerProThrSerGluGluPheAspTrp-GlyThrThrAsnCysT 80
        |||||||||||||||||||||||| ::||| |||||||| ||||||||||||||||
Db  200 TCAAACTGGAAAAAAGTCCCACTTCTGAATACTGTTTGGACTGGGGGACCACCAAATTGGA 259

Qy   80 hrValGlyAspLeuValAspLeuLeuGluAsnGluPheAlaProAlaSerLeuI 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  260 CAGTTGGTGATCTTGTGGGATCTTTTGATCCAAAATGAATT-TTTGCTCCTGGAGTCTTT 318

Qy  100 euLeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrV 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  319 TGCTCCCAGATGCTGTTCCCAA-ACTGCTAATACACTACCTTCTAAGAAAGCTATAACAG 377

Qy  120 alGlnGlnLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnA 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  378 TTCAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGGACATTGATGACACCTGTGCANA 437

Qy  140 snLeuGluGlnSerTyrMetProProAspSerSerProGluAsnLysSerLeuGluV 160
        ||||||||||||||||||||||||||||||||||||||||||||||::|||||||||
Db  438 ATCTTGAACAAAGCTATATGCCACCTGACTCCTCAAGTCCANAAAATAAAAGTTTAAAG 497

Qy  160 al 160
        ||
Db  498 TT 499
```

```
;  SEQ ID NO 10
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
US-09-966-451-10

Alignment Scores:
Pred. No.:
Score:                              2.44e-54
Percent Similarity:    Length:      590.00
Best Local Similarity: Matches:     26.94%
Query Match:           Conservative: 26.31%
DB:                    Mismatches:  26.31%
                       Indels:      24.73%
                       Gaps:        11

                       31000
                       166
                       4
                       6
                       455
                       2
```

```
US-10-001-254-16 (1-460) x US-09-966-451-10 (1-31000)

Qy    53  IleArgArgPheGluAlaLeuLeuLeuGlnThrGlyLysSerProThrSerGlyLeuLeuPhe  72
          ::::|||::|||||||||||||||:|||||||||||:||||::|||||||||||||||||
Db  12963  TTAAGGAGGATTGGAAAGACAATTACTTCAAACTGGGAAAAAGTCCCACTTCTGAATTACTGTTT  13022

Qy    73  AspTrpGlyThrThrAsnCysThrValGlyAspLeuLeuIleGlnAsnGlu  92
          ||:|||||||||||||||||||:||||||||||:|||||||||
Db  13023  GACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGAATCTTTTGATCCAAAATGAA  13082

Qy    93  PhePheAlaProAlaSerLeuLeuLeuPro------------------------  102
          ||||||||||||||||||||:|||
Db  13083  TTTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGGTAAACTGATTGTGACCAGGGTGTCCACA  13142

Qy   102  ------------------------------------------------------  102
Db  13143  ATTAGGGTGGAAAGACAAATGGCACGAAAATATAAAGTTTCTTCTTACTCTTCCTTTTTTTC  13202

Qy   102  ------------------------------------------------------  102
Db  13203  TCATAGTAGATGAAGCTTACACATTTGAGAGTCCCTTTCTTCAGCACTCCTCAACTCTTTA  13262

Qy   102  ------------------------------------------------------  102
Db  13263  AAAAGCAGCACAGACAAAAGACACACTGGGACTCTCGCTGCTAAGGTGATAGAAGCTTCGTA  13322

Qy   102  ------------------------------------------------------  102
Db  13323  AGAGTTAGAATAGTTTTGTGCCAACAGGAGAGTTTAGAAGGAAAGACTTCATACTTCTGGCT  13382

Qy   102  ------------------------------------------------------  102
Db  13383  TAGGCTGTAAGAAAGTAATTAAATTTATAATTTTGAGTCTTCTTCCTTTTTTCATCTTCAACTTCTA  13442

Qy   102  ------------------------------------------------------  102
Db  13443  CCCTGATGGGACTCTATAATCATATAATTTTAAAAAAATTGTATTCTGATGTGTAGTGTCTA  13502

Qy   102  ------------------------------------------------------  102
Db  13503  GTTGGCTTCCTTATAAGCTTTTCTTTTTTCCTTTGATGCCTTTCACCAGTGGTTCTCA  13562

Qy   102  ------------------------------------------------------  102
Db  13563  ACCTTTTATCATGCATAAGAATCACTTAGGTTCGGTTAAACATGTAGTTTCCTAGGGTCT  13622

Qy   102  ------------------------------------------------------  102
Db  13623  GTTCCAAAATCTATATGTAAATCAGAATTCTGGAATTGGGTTTAAGGTTCTATATTCTTAAAACAG  13682

Qy   102  ------------------------------------------------------  102
Db  13683  GTGCTTCAGTGCCTCTGATATAGGTGGGCTCATGGAATCACTGTTTGAAAACACTGCCTTAC  13742

Qy   102  ------------------------------------------------------  102
Db  13743  TCTCTTTACCCATCTTCATTATAACTTACAGAGATTCTTACTAGGCAGCTTCTTGTGTGTGC  13802
```

```
Qy   102  ------------------------------------------------------  102
Db  13803  TGTGAGAATATGAGACCAACCTGTAGAAACTGGAATGATGATTAAATGAACCAAGTTTCTA  13862

Qy   102  ------------------------------------------------------  102
Db  13863  GTTTAACTTTTTCACAACCACTTTTTCTTACTGAAAAACCACTTGTATCTTACTTCATTT  13922

Qy   103  ---AspAlaValProLysThrAlaAsnThrLeuProSerLysSerGluAlaIleThrValGl  121
          :|||||||||||||||||||||||||||||||||||||||||||||||||:|
Db  13923  GTTAGATGCGTGTTCCCAAAACTGCTAATACACTACCTTCTAAAAGAAGCTATAAACAGTTCA  13982

Qy   121  nGlnLysGlnMetProPheCysAspGlyAspArgThrLeuMetThrProValGlnAsnLe  141
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13983  GCAAAAAACAGATGCCTTTTCTGCTGACAAAGACAGGACAATTGATGACACCTGTGCAGAATCT  14042

Qy   141  uGluGlnSerTyrMetProProAspSerSerProGluAsnLysSerLeuGluGluValSe  161
          |||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db  14043  TGAACAAAGCTATATGCCACCTGACTCCTCCAGTCCAGAAAATAAAAGTTTAGAAGAAGTTAG  14102

Qy   161  rAspThrArg-----------------------------------------------  164
          ||:|||||||
Db  14103  TGATACACGTAAGTAACACATTTTCAGTGCTTTCCACTAGGGATTTGTTCATTAAGACTACCA  14162

Qy   164  ------------------------------------------------------  164
Db  14163  GTGCTTTAAAAGAAAGCTCTTGCTCTTTGTTGTGCAGCAATCACAGGACACACTGGCCA  14222

Qy   164  ------------------------------------------------------  164
Db  14223  TAGCTCTTTTGTTGAGTTGTTTCCTCCTGATATATTAAGAACCATCTTCATTGTATTAATC  14282

Qy   164  ------------------------------------------------------  164
Db  14283  AGTGATTAGAGGCAATAGGACATGCAAACAACCAGAGAGCTATGAAAAAAAATAAGTAA  14342

Qy   164  ------------------------------------------------------  164
Db  14343  ATATTTACATAACTTGGAAAGGTCACTTTTTAAAATATAAAATATGTGACTAGAGTTTTTGGG  14402

Qy   164  ------------------------------------------------------  164
Db  14403  TAGGTAGACTAGACCCACTCAAAATGTGTGACTAGAATTTGGGGTGGGGTGACTACAGTGTT  14462

Qy   164  ------------------------------------------------------  164
Db  14463  CAGAGGGTAGGATCACCAAACAGAGAGTTCTCAAGAAAAAACTTATAATTGTATTTTTAGAAA  14522

Qy   164  ------------------------------------------------------  164
Db  14523  TGGGTTTTATCTTCTCTCATCTTGTCTTAATTCAAAAATCATTAAAATATTTGCATATATGTGAA  14582

Qy   164  ------------------------------------------------------  164
Db  14583  ATCTTCAAAATGAGAAAATATTATTAATTTATAAAACGAATTTTTAAAAAATTTAAGCACATGTTT  14642

Qy   165  ------------------PheHisSerPheSerPheTyrGluLeuLeuAsnValThrAs  178
          ||||||||||||||||||||||||||||||||||||||
Db  14643  TTCTTATTTTGACATAGGTTTTCACAGTTTTTATCATTTTATGAATTGAAGAAGTGTCACAAA  14702

Qy   178  nAsnPheAspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGl  198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  14703  TAACTTTGATGGAACGACCCATTTCTCGTTGGTGGTAATAAAATGGGAGGGGGAGGATTTGG  14762

Qy   198  yValValTyrLysGlyTyrTyrValAsnAsnThrThrValAlaValLysLysLeuAlaAlaMe  218
          |||||||||||:|||||||||||||||||||||||||||||||||:::
Db  14763  AGTTGTATATAAAAGGCTACGTAAAAATAAACAACTGTGGCACTGGAAGAAAGCTTGCAGCAGT  14822

Qy   218  tValAspIleThrThrGluGluLysGln  228
          :::|||||||||||:::
Db  14823  AAGTTATATTTTCAGGAATAAAAAAGAAAGAG  14853
```

```
RESULT 10
```

```
; US-10-001-254-5
; Sequence 5, Application US/10001254
; Publication No. US20030049702A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sug Hyung
; APPLICANT: Roth, Wilfred
; APPLICANT: Stenner-Liewen, Frank
; TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins
; FILE REFERENCE: P-LJ 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 5
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(294)

US-10-001-254-5

Alignment Scores:
Pred. No.:              3.77e-49
Score:                  517.00
Percent Similarity:     100.00%
Best Local Similarity:  100.00%
Query Match:            21.67%
DB:                     15

                        Length:         294
                        Matches:        98
                        Conservative:   0
                        Mismatches:     0
                        Indels:         0
                        Gaps:           0

US-10-001-254-16 (1-460) x US-10-001-254-5 (1-294)

Qy   9 ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgGlyLysLeuSerAspPheIleAspPro 28
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACATATGTGCCGCTGCCTCAAGTTGGACTAATAGGAAGCTTGTCAGATTTTATTGATTCCT 60

Qy  29 GlnGluGlyTyrPheLysSerGlyAspAspArgTyr 48
        ||||||||||||||||||||||||||||||||||||
Db  61 CAAGAAGGATCGAAGAAAGTTAGCTGTAGCTATTAAAAAAACCATTCTGGTTGATGATAGGATAC 120

Qy  49 AsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrArgGlyLysSerProThrSer 68
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AATCAGTTTCACATAAGGAGAATTGAAGCATTACTTCAAACTCGGTGATGATGAGAGATCC 180

Qy  69 GluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuAspLeu 88
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTACTTGATGACTTTG 240

Qy  89 IleGlnAsnGluPheHisAlaProAlaSerLeuLeuLeuLeuProAspAlaValPro 106
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 ATCCAAAATGAATTTTTTGCTCCTCCGGAGTCTTTTCCTCCCAGATGCTGTTCCC 294
```

```
; US-09-880-107-2222
; Sequence 2222, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
;
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 2222
; LENGTH: 3590
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L76191

US-09-880-107-2222

Alignment Scores:
Pred. No.:              2.43e-47
Score:                  515.50
Percent Similarity:     46.14%
Best Local Similarity:  29.38%
Query Match:            21.61%
DB:                     10

                        Length:         3590
                        Matches:        156
                        Conservative:   89
                        Mismatches:     192
                        Indels:         94
                        Gaps:           19

US-10-001-254-16 (1-460) x US-09-880-107-2222 (1-3590)

Qy   4 ProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgGlyLysLeuSer 23
        ||  ::::||||  ::::||||  ||||||||||::||| :::::||||||||||
Db 115 CCCCCGGGGCCCCAGCACTTCTTCTTCTACCGAGGTGCCGCCTTCGCCGTTCTACAA 174

Qy  24 AspPheIleAspProGlnGluGlyTyrPheLysSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrArgGlyLysSerProSer 43
        |||  ::: |||  :::::::||  ||||::::|::||| ::::||  ||
Db 175 AGT-----GATGGACGGCCCCTGGAGCCCCGGACTG--GTGCCAGTTCGCCGCCT--- 222

Qy  44 GlyAspAspArg-TyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeuLeuGlnThrGl 63
        |||||  ::: |||  :::::||||| :::||  ::::||
Db 223 ----GATCGTGCGCGAACCAGACCGAGCTGCGG--- 264

Qy  63 yLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAs 83
        ||  :::  ::::||  ::||| ::||||||:: ||||  :::||||||
Db 265 CGGGCACGCACGGCCAGGCGTCCTGTGGCCCTGGATCAACCCGACAGCCCGTGTGGCCGA 324

Qy  83 pLeuValAspLeuLeu----IleGlnAsnGluPh 93
        |||||  :::||| :::||||||
Db 325 CCTCGTGCACATCCTCACGCAACCTGCGAGCTGTCCGGGGGGGGGACAATCATCACAGCCTG 384

Qy  93 ePheAlaProAlaSerLeuLeuLeuLeuProAspAlaValProLysThrAlaAsnThrLeuPr 113
        ||:::::: ||  ::::||||||
Db 385 GCACCCTCCCCGCCCCCGCTTCCGTCCCCAGGCACCACTGCCCCGAGGCGCCAGCAGCATCCC 444

Qy 113 oSer-----LysGluAlaIleThrValGlnGlnLysGlnMetProPheCysAspLysAspAr 132
        ::: |||
Db 445 TGCACCCGCCCAGGCCCCGAGGCCTGGAGCCCCGGAAGTTGCCA---- 498

Qy 132 gThrLeuMetThrPro---- 137
        |||
Db 499 CACCTTTCCTCTCCCCAGCTTTTCCAGGCTCCCAGCCCCATTCAGGGCCTGAGCTCGGCCT 558

Qy 138 -ValGlnAsnLeuGlnSerTyrMetProProAspSerSerProGluAsnLysSe 157
        ||||||  :::||||||
Db 559 GGTTCCAAGCCCCTGCTTCCCGTGGGCCTTCCACCGCCCCTTCTTCTACCAA 618

Qy 157 r----------------LeuGluValSerAspThrArgPheHisSerPh 168
        ||
Db 619 GCCAGGCCCCAGAGAGCTCAGTGTCCCTTCCCTGCAGGGAGCCGCGCCCCTCTCCGTTTGCTG 678

Qy 168 eSerPheTyrGluLeuLysAsnValThrAsnAsnPheAspGluArgProIleSerValGl 188
        : ||||:::: ||||||
Db 679 GCCCCCTCGTGAGATTTCCCGGGGGGCACCCCACAACTTCTCCGGAGGAG---------- 724

Qy 188 yGlyAsnLysMetGlyGluGlyTyrPheGlyValValThrTyrLysGlyTyrValValAsnTh 208
        ||||||||:::: |||  ::::||||||
Db 725 ------CTCAAGATCGGGGAAGGGGGTGGCTTTGGGGTGGTGACGTACAAAGGGTATGTGGTGAATGAGGAACAC 780

Qy 208 rThrValAlaValLysLysLeuAlaAlaMetValAspAspIleThrThrGluGluLeuLysGl 228
        ||||  :::||||||||||
Db ...
```

```
;
; SEQ ID NO 4
; NUMBER OF SEQ ID NOS: 142
; CURRENT FILING DATE: 2002-06-10
; CURRENT APPLICATION NUMBER: US/10/167,034
; FILE REFERENCE: PTS-0003
; TITLE OF INVENTION: ANTISENSE MODULATION OF IL-1 RECEPTOR-ASSOCIATED KINASE-1 EXPRESS
; APPLICANT: Kenneth W. Dobie
; APPLICANT: Susan M. Freier
; APPLICANT: Brenda F. Baker
; GENERAL INFORMATION:
; Publication No. US20030228690A1
; Sequence 4, Application US/10167034
US-10-167-034-4
RESULT 12
```

```
;
; DB:               US-10-167-034-4
; LOCATION:         (80)...(2218)
; NAME/KEY:         CDS
; FEATURE:
; FEATURE:
; ORGANISM:         H. sapiens
; TYPE:             DNA
; LENGTH:           3590
```

```
Alignment Scores:
Pred. No.:              2.43e-47
Score:                  515.50          Length:         3590
Percent Similarity:     46.14%          Matches:        156
Best Local Similarity:  29.38%          Conservative:   89
Query Match:            21.61%          Mismatches:     192
DB:                     12              Indels:         94
                                        Gaps:           19
```

US-10-001-254-16 (1-460) x US-10-167-034-4 (1-3590)

RESULT 13
US-10-101-510-88

US-10-101-510-88

Alignment Scores:

Pred. No.:	2.43e-47	Length: 3590
Score:	515.50	Matches: 156
Percent Similarity:	46.14%	Conservative: 89
Best Local Similarity:	29.38%	Mismatches: 192
Query Match:	21.61%	Indels: 94
DB:	13	Gaps: 19

US-10-001-254-16 (1-460) x US-10-101-510-88 (1-3590)

(Amino acid / nucleotide sequence alignment blocks follow, arranged in paired Qy (query) and Db (database) rows with alignment bars; individual residue-level content is rendered in the original figure.)

SEQ ID NO 1003
LENGTH: 3590
TYPE: DNA
ORGANISM: Homo sapiens
US-09-873-367C-1003

Alignment Scores:

Pred. No.:	2.43e-47	Length: 3590
Score:	515.50	Matches: 156
Percent Similarity:	46.14%	Conservative: 89
Best Local Similarity:	29.38%	Mismatches: 192
Query Match:	21.61%	Indels: 94
DB:	13	Gaps: 19

US-10-001-254-16 (1-460) x US-09-873-367C-1003 (1-3590)

```
Qy     4  ProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSer  23
          |||||||||||||||||||||||||             :::|||
Db   115  CCCCCGGCGCCCCAGCACTTCTTGTACGAGGTTGCCCCCGCCCTGGGTCATGTGCCGGTCTACAA  174

Qy    24  AspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleIleLysLysProSer  43
          ||||||||                  ::::||||||||||
Db   175  AGT------GATGGACGCCCTGGGACCCCCGACTG---GTGCCAGTTCGCCGCCT---  222

Qy    44  GlyAspAspArg-TyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGl  63
          |||||||||         ::::|||
Db   223  ------GATCGTGCGGCGACCAGACCGAGCTGCGG------CTGTGCGAGCGGCTC  264

Qy    63  TyrLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAs  83
          |||||                   ::::||||||
Db   265  CGGGGCAGCGGCGACCGAGCCTGAGCCGTCGTGTGGGCCCCTGGATCAACCGGCAACGCGCTGTGGGCCGGA  324

Qy    83  PLeuValAspLeuLeu-------------------IleGlnAsnGluPh  93
          :::|||                           |||||
Db   325  CCTCGTGCACATCCTCCACGCACCTGCAGCTGCTCCGTGCGGGACATCACAGGCTG  384

Qy   132  GThrLeuMetThrPro-------------------  137
          :::|||
Db   445  TGCACCCGCGCCGAGGGCCTGGAGCCCGGGCCCGGAAGTGTCCA------TCCTCAGCCTC  498

Qy   168  eSerSerPheTyrGluLeuLysAsnValThrAsnAsnPheAspGluArgProIleSerValGl  188
          ::::|||
Db   619  GCCAGCGCCCAGAGAGACTCACTGTCCCTCTGGCAGGGACGCCGCCCCCTCCGTTTGCTG  678

Qy   188  -ValGlnAsnLeuGluGlnSerTyrMetProProAspSerSerProGluAsnLysSe  208
          |||           :::|||
Db   679  CACCTTCCCTCCCCCAGCTTTCCAGGGGTCCCCAGACCCAATTCAGGGCCCTGAGCTCGGGCCT  558

Qy   208  rThrValAlaValLysLysLeuLeuAlaAlaMetValAspIleThrThrGluGluLysGl  228
          |||||              :::|||
Db   781  GGTGTATGCGTGTGAAGAGGGTTGAAGGAGGAACGCTGACCCTGGACTCCAGTGAAGCA  840

Qy   228  nGlnPheAspGlnIleLysValMetAlaLysCysGlnHisGluAsnLeuValGluLe  248
          |||:|||
Db   841  GAGCTTCCTGACCGAGGTGGAGCAGCTGTCCAGGTTTCGTCACCCAAACATTGTGGACTT  900

Qy   248  uLeuGlyPheSerSerAspGlyAspAspLeuCysLeuValTyrValTyrMetProAsnGl  268
          |||||
Db   959  GGTGTACGAGCAGGGTCTTAGAGAAGCTGCAGGCA  1648
```

RESULT 15
US-09-938-842A-1673
Sequence 1673, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
    APPLICANT: Harper, Jeff
    APPLICANT: Kreps, Joel
    APPLICANT: Wang, Xun
    APPLICANT: Zhu, Tong
    TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
    TITLE OF INVENTION: SAME, AND METHODS OF USE
    FILE REFERENCE: SCRIP1300-3
    CURRENT APPLICATION NUMBER: US/09/938,842A
    CURRENT FILING DATE: 2001-08-24
    PRIOR APPLICATION NUMBER: US 60/227,866
    PRIOR FILING DATE: 2000-08-24
    PRIOR APPLICATION NUMBER: US 60/264,647
    PRIOR FILING DATE: 2001-01-16
    PRIOR APPLICATION NUMBER: US 60/300,111
    PRIOR FILING DATE: 2001-06-22
    NUMBER OF SEQ ID NOS: 5379
    SEQ ID NO 1673

;
; LENGTH: 2361
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-1673

Alignment Scores:
    Pred. No.:
    Score: 508.50   Length: 2361
    Percent Similarity: 55.79%   Matches: 124
    Percent Identity: 36.80%   Conservative: 64
    Best Local Similarity: 21.31%   Mismatches: 118
    Query Match: 10   Indels: 31
    DB: Gaps: 9

US-10-001-254-16 (1-460) x US-09-938-842A-1673 (1-2361)

```
Db   2098  ACGGGAGAT------ATTGGAAATATTGTTGATCCAAACCTTCACGGAGCTTACGACGTC 2151

Qy    424  ThrSerValGluAlaMetTyrSerValAlaSerGlnCysLeuHisGluLysLysAsnLys 443
                 ||||||||            |||:::::::
Db   2152  GGTTCTGTCTTGGAAGGCTATTGAACTAGCTATGTCATGTGTGAATATTTCTTCGGCAAGA 2211

Qy    444  ArgProAspIleLysLysValGlnLeuLeuGlnGluMetThrAlaSer 460
                 |||||||||            |||::::|||            |||
Db   2212  AGACCAAGACATGTCTCAAGTTGTTAGCGATCTTAAAGAGTGTGATATCT 2262
```

Search completed: January 18, 2004, 09:06:01
Job time : 721.136 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 18, 2004, 01:11:43 ; Search time 4491.9 Seconds
                 (without alignments)
                 2488.940 Million cell updates/sec

Title: US-10-001-254-16
Perfect score: 2386
Sequence: 1 MNKPITPSTYVRCLNVGLIR..........KNKRPDIKKVQQLLQEMTAS 460

Scoring table: BLOSUM62
                Xgapop 10.0 ,   Xgapext 0.5
                Ygapop 10.0 ,   Ygapext 0.5
                Fgapop 6.0 ,   Fgapext 7.0
                Delop   6.0 ,   Delext   7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Maximum DB seq length: 200000000
Minimum DB seq length: 0

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152424_19723/app.query.fasta_1.1109
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US10001254 @CGN 1_1_4382 @runat_16012004_152424_19723 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
                1: em_estba:*
                2: em_esthum:*
                3: em_estin:*
                4: em_estmu:*
                5: em_estov:*
                6: em_estpl:*
                7: em_estro:*
                8: em_htc:*
                9: em_htc1:*
                10: gb_est1:*
                11: gb_htc:*
                12: gb_est3:*
                13: gb_est4:*
                14: gb_est5:*
                15: em_estom:*
                16: em_estfun:*
                17: em_est5:*
                18: em_gss_hum:*
                19: em_gss_inv:*
                20: em_gss_pln:*
                21: em_gss_vrt:*
                22: em_gss_fun:*
                23: em_gss_mam:*
                24: em_gss_mus:*
                25: em_gss_pro:*
                26: em_gss_rod:*
                27: em_gss_phg:*
                28: em_gss_vrl:*
                29: gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB ID	Description
1	2017	84.5	2481	11	AK028837	AK028837 Mus muscu
2	2004	84.0	2810	11	AK029028	AK029028 Mus muscu
3	1094.5	45.9	859	13	BG164491	BG164491 602342026
4	1004.5	42.1	682	10	BF694134	BF694134 602082746
5	946	39.6	718	13	BQ780493	BQ780493 UI-R-FF0-
6	946	39.6	1161	11	AK020397	AK020397 UI-R-FF0-
7	944	39.6	702	13	BQ782157	BQ782157 UI-R-FF0-
8	935.5	39.2	811	10	BG616438	BG616438 602642772
9	874	36.6	769	13	BU441365	BU441365 603208981
10	856	35.9	719	10	BF696981	BF696981 602130160
11	804	33.7	962	10	BU225249	BU225249 603947374
12	774.5	32.5	541	10	BG691069	BG691069 340084 BA
13	755	31.6	676	10	BB613167	BB613167 602066996
14	752	31.5	478	12	BM431425	BM431425 1Duo1d005
15	749.5	31.4	858	10	BF687921	BF687921 602066996
16	675	28.3	635	13	BQ205658	BQ205658 UI-R-EPo-
17	665	27.9	424	11	BF758197	BF758197 RC4-CT052
18	662.5	27.8	624	13	BU479576	BU479576 603472624
19	640.5	26.8	508	10	BE479760	BE479760 164577 BA
20	639	26.8	663	10	BY726858	BY726858
21	632	26.5	557	14	CB475590	CB475590 jns109_F0
22	631	26.4	637	10	BB613447	BB613447 60190613
23	629	26.4	576	10	BF238344	BF238344 60190613
24	627	26.3	638	14	BY721552	BY721552 60378255
25	623	26.1	613	13	BU316755	BU316755 603850730
26	612	25.6	598	13	BQ552228	BQ552228 H401AC09-
27	597	25.0	508	13	BX283594	BX283594 BX283594
28	588	24.6	1090	12	BM918155	BM918155 AGENCOURT
29	587	24.6	610	10	BB660378	BB660378 UI-R-FF0-
30	580	24.3	499	13	BU759126	BU759126 UI-R-FF0-
31	577.5	24.2	852	13	BU209111	BU209111 603950834
32	568.5	23.8	710	14	CA054774	CA054774 ssalrgb51
33	563	23.6	335	10	BU430802	BU430802 UI-HF-BN0
34	557	23.3	402	10	BE482619	BE482619 168463 BA
35	556	23.3	823	13	BU246150	BU246150 603782255
36	551.5	23.1	555	13	BQ617704	BQ617704 faa67c09.
37	549	23.0	575	4	BX522921	BX522921 R2FD Mus
38	528	22.1	503	14	CA538859	CA538859 C0272R01-
39	523	21.9	851	13	BU255115	BU255115 603746394
40	522	21.9	773	13	BU240307	BU240307 603321515
41	511.5	21.4	629	9	AL647125	AL647125 AL647125
42	504	21.1	804	14	CB659049	CB659049 OSJNE15K
43	488.5	20.5	505	10	BE482323	BE482323 168064 BA
44	487.5	20.4	2675	11	AK045958	AK045958 Mus muscu
45	486.5	20.4	1434	9	AI967314	AI967314 Ljirnpest

ALIGNMENTS

RESULT 1
AK028837
LOCUS   AK028837   2481 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION   Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
               library, clone:4732460109 product:interleukin-1 receptor-associated
               kinase 4 [Mus musculus], full insert sequence.
ACCESSION   AK028837
VERSION   AK028837.1 GI:26324783
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED   10349636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
             Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
             Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
             Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
             Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
             Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
             Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
             Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
             Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
             Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
             Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
             Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
             Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
             Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
             Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
             Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
             Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
             Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
             and Hayashizaki,Y.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409 (6821), 685-690 (2001)
MEDLINE   21085660
PUBMED   11217851
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6   (bases 1 to 2481)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome

COMMENT   Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
             CDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in Riken.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.go.jp/
             URL:http://fantom.gsc.riken.go.jp/.

FEATURES   Location/Qualifiers
     source   1..2481
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:4732460I09"
                     /db_xref="taxon:10090"
                     /clone="4732460I09"
                     /tissue_type="skin"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     CDS   116..1495
                     /note="unnamed protein product; interleukin-1
                     receptor-associated kinase 4 [Mus musculus]
                     putative"
                     /codon_start=1
                     /protein_id="BAC26146.1"
                     /db_xref="GI:26324784"
                     /translation="MNKPLTPGTYIRNLNVGIIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLLVQIELFAPATLLLP
                     DAVPQTVKSLPREAATVAQTHGPCQKRDRTSVMPMPKLEHSCEPPDSSSPDRRSVES
                     SDTRFHSFSFHELKSITNNFDEQPASAGGNRMGEGGFGVVYKGCVNNTIVAVKKLGAM
                     VEISTEELKQOFDQEIKVMATCQHENLVELLGFSSDSSDNLCLVYAYMPNGSLLDRLSC
                     LDGTPPLSWHTRCKVAQGTANGIFRLHENHHIRDIKSANILLDKDFTAKISDEGLAR
                     ASARLAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLELITGLAAVDENRE
                     PQLLIDIKEEIEDEEKTIEDYTDEKMSDADPASVEAMYSAASQCLHEKKNRRPDIAKV
                     QQLLQEMSA"

BASE COUNT   650 a   592 c   634 g   605 t

ORIGIN

US-10-001-254-16 (1-460) x AK028837 (1-2481)

Alignment Scores:
Pred. No.:   1.18e-221
Score: 2017.00
Percent Similarity: 90.85%
Best Local Similarity: 83.88%
Query Match: 84.53%

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
Db   116 ATGAACAAGCCGTTCACCACCATCGACATACATCAGCAACCTTAATGTGGGGATCCTTAGG 175

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys 40
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
Db   176 AAGCTGTCGGATTTTATTGATCCTCAAGAGGGGTGGAAGAAATTAGCAGTAGCTATCAAA 235

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
Db   236 AAGCCGGGTCCGGACGACAGATACAATCAGTTCCATATATAAGGAGATTCGAAGCCTTACTT 295

Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr 80
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |
Db   296 CAGACGGGAAGAGCCCCACCTGTGAACTGCTGTTTGACTGGGGACCACAGAACTGCACA 355

Qy   81 ValGlyAspLeuValAspLeuLeuLeuValGlnIleGluLeuPheAlaProAlaSerLeuLeu 100
          ||| ||| | |::|| :||| :|| :::||||| || || |||::|||||| |||::|||||| |
Db   356 GTTGGCGACCTTGTGGATCTACTGGTCCAGATTGAGCTGTTTGCCCCCGCCACTCTCCTG 415

RESULT 2

```
Qy   101 LeuProAspAlaValProLysThrAlaAlaAsnThrLeuProSerLysGluAlaIleLeuThrVal 120
             |||||          ::|||||
Db   416 CTGCCCGGATGCCGTTCCCAAACGTCAAAAACGCTGCCTCCTAGAGAAGCGGCCAACAGTG 475

Qy   121 GlnGlnLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn 140
             ::|||
Db   476 GCACAAACACACGGGCCTTGTCAGGAAAAAGACACAGCAACATCCGTAATGCCTTATGCCGAAG 535

Qy   141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLeuSerLeuGluVal 160
             |||
Db   536 CTAGAACACAGCTGCGAGCACCTCCTCAAGCCACAGCAACAGAAGTTGAGTCC 595

Qy   161 SerAspThrArgPheHisSerPheTyrGlyLeuLysSerAsnValThrAsnAspAsnPhe 180
             |||||||||:::::
Db   596 AGCGACACTCGGTTCCACAGCTTCTCGGTTCCATGAACTAGAGAGCATCACAAACACTTC 655

Qy   181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyGlyGlyGlyPheGlyValVal 200
             :::|||
Db   656 GACGAGCAAACCCGCGTCTGCCGGTGGCAACCGGATGGGGAGGGGGGGATTTGGAGTGGTG 715

Qy   201 TyrLysGlyTyrValAsnAsnThrThrValAlaValLysLysLeuAlaAlaMetValAsp 220
             |||||||||
Db   716 TACAAGGGCTGTGTGAACAACACCACGTGGGTGGGAAGAAGCTCGGAGCCATGGTTGAA 775

Qy   221 IleThrThrGluGlnLeuLeuLysGlnLeuAsnPheHisAspGlnIleLeuValMetAlaLysCys 240
             |||::|||||
Db   776 ATCACGACTGAGAACTAAAGCAACACAGTTTGATCAAGAAATTAAAGTAATGGCAACGTGT 835

Qy   241 GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspLeuCysLeu 260
             ||||||:::
Db   836 CAGCACGAGAACCTGTGGGAACTGCTCGGCTTCTCCAGCGACGCACAACCTGTCCTTA 895

Qy   261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr 280
             ::: |||
Db   896 GTGTATGCTTACATGCCCAACGGGTCCTTGCTGGACAGACGTCTGCCTGGATGGTACA 955

Qy   281 ProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsn 300
             |||:::
Db   956 CCACCCGCTTCCTGGCACATGAGGTGCAAGGTGGCTCAGGGAGCCGCAAATGGCATCAGG 1015

Qy   301 PheLeuHisGluAsnHisHisHisArgAspIleLysSerAlaAsnIleLeuLeuAsp 320
             |||||
Db  1016 TTTCTGCCATGAAAATCATCACATTCATAGAGTATTTAAAAGTGCAAATATCTTACTTAGAC 1075

Qy   321 GluIleAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAla 340
             :: |||
Db  1076 AAAGACATTTACTGCCAAAATATCTGACTTTGGGCTTGCACGGGCTTCGGCAAGGCTAGCG 1135

Qy   341 GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu 360
             |||:::
Db  1136 CAGACGGTCATGACCAGCCGAATCGTGGGCACACGGCCTTACATGGCCCGAAGTTTG 1195

Qy   361 ArgGlyGlyIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIle 380
             |||:::
Db  1196 CGGGGAGGAAATAAACACCCAAATCTGACATCTACAGCTTCGGCGTTGTTCGTTGAGCTG 1255

Qy   381 IleThrGlyLeuProAlaValAspGluHisArgGluProGlnLeuLeuLeuAspIleLys 400
             |||||
Db  1256 ATAACCGGGTCTGCCGGCCTGGATGAAAACCTGCAACTACTGCTGGATAATAAA 1315

Qy   401 GluGluIleLeuAspGluGluGlnLysThrIleGluAspTyrIleAspLysLysMetAsnAsp 420
             ||||| ||||
Db  1316 GAAGAGATTGAAGATGAAGAGGAAGACGATTGAAGATTACACGGATCAGAACAGATGACCGAT 1375

Qy   421 AlaAspSerThrSerValGlnAlaMetTyrSerValAlaSerGlnCysLeuHisGluLys 440
             |||||
Db  1376 GCCGGACCCTGCTTCCGTGGTGGGAGCAAGTACTCTGCTGCTCAGCCAGTGTCCGCATGAGAAG 1435

Qy   441 LysAsnLysArgProAspIleLysGlnIleLeuLeuGlnLeuMetThrAla 459
             |||||
Db  1436 AAAAAACAGACGGCCAGACAATTGCAAAGGTTCAACACTGCTCAGAGAGATCTCGCT 1492
```

LOCUS   AK029028   2810 bp   mRNA   linear   HTC 05-DEC-2002
DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
            library, clone:4732482P03 product:interleukin-1 receptor-associated
            kinase 4 [Mus musculus], full insert sequence.
ACCESSION   AK029028
VERSION   AK029028.1 GI:26081132
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   2
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 2810)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,

Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

TITLE   Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
         Physical and Chemical Research (RIKEN), Laboratory for Genome
         Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
         RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
         Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
         Tel:81-45-503-9222,
         Fax:81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome
         Encyclopedia Project of Genome Exploration Research Group in Riken
         Genomic Sciences Center and Genome Science Laboratory in RIKEN.
         Division of Experimental Animal Research in Riken contributed to
         prepare mouse tissues.
         Please visit our web site for further details.
         URL:http://genome.gsc.riken.go.jp/
         URL:http://fantom.gsc.riken.go.jp/.

FEATURES   Location/Qualifiers
source   1..2810
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4732482P03"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /tissue_type="skin"
                     /dev_stage="10 days neonate"
misc_feature   176..1556
                     /note="interleukin-1 receptor-associated kinase 4 [Mus
                     musculus]
                     putative"

BASE COUNT   738 a   677 c   715 g   680 t
ORIGIN

US-10-001-254-16 (1-460) x AK029028 (1-2810)

Alignment Scores:
Pred. No.:
Score: 4.47e-220
                         204.00
Percent Similarity: 90.65%
Percent Identity: 83.70%
Best Local Similarity: 83.99%
Query Match: 11
DB:

Length: 2810
Matches: 385
Conservative: 32
Mismatches: 42
Indels: 1
Gaps: 0

RESULT 3

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||:::||
Db 176 ATGAACAAGCCGTTGACACCATCGACATACCAACACTTAATGTGGGAGGATCCTTAGG 235

Qy   21 LysLeuSerAspPheIleAspProGlnGlyThrArgLysGlySerLeuAlaValAlaLeuLys 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 236 AAGCTGTCCGGATTTTATTGATCCTCAAGAAGGTGGAAGAAATTAGCAGTGAGCTATCAAA 295

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
        |||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db 296 AAGCCGTCCGGCGACGACAGATACAATCAGTTCCATATAAAGGAGATTCGAAGCCTTACTT 355

Qy   61 GlnThrGlyLysSerProThrSerGluValAlaMetThrSerValAlaSerLeuLeu 80
        |||||||||||||||||||||||||||:||||||||||||||||||||||||
Db 356 CAGACCGGGAAGAGCCCACCCTGTGAAGTAGCCATGACTTCGGTGGCCAGTCTGCTA 415

Qy   81 ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPhePheAlaProAlaSerLeuLeu 100
        |||||||||||||:||||||||||||||||:||||:|||||||||||||||||||||||
Db 416 GTTGGCGACCTTGTGGATCTACTGCTTCAGATTGAGCTGTTTGCCCCGCCACTCTCCTG 475

Qy 101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerTyrGluAlaIleThrVal 120
        ||||||||||||||||:::|||||||||||||||||||::::::|||||||||||||||
Db 476 CTGCCCGGATGCGTTCCCAAAACCCTCAAAACGGCTGCCTTCCTTAGAGAAGGCCAACAGTG 535

Qy 121 GlnGlnLysGlnMetProPheCysAspIleAspArgThrLeuMetThrProValGlnAsn 140
        ||| :::|||||||||||||||||||:::||||||||
Db 536 GCACAAAACACACGGCCCTTGTCAGGAAAGGAACAGGACAATCCGTAAATGCCTATGCGAAG 595

Qy 141 LeuGluGlnSerTyrMetProProAspSerSerProGluAsnLysSerLeuGluVal 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 596 CTAGAACACAGACGCTGCGAGCCACCCGACTCCTCAAGCCGAACAGAAGAGTGTAGAGTCC 655

Qy 161 SerAspThrArgPheHisSerPheSerPheTyrGluLysAsnValThrAsnAsnPhe 180
        |||||||||||||||||||||||||:::||||||||:::||||||||||||||||||
Db 656 AGCGACACTCGGTTCCACAGCTTCTCGTTCTACGAGAAGAACGTAACTAACAACTTC 715

Qy 181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyPheGlyValVal 200
        :::||||||||||||||||||||||||||||||||||||::|||||||:||||||||
Db 716 GACGAGCGACCCGGGTCTCCGGGTGGCAACCGGATGGGAGAGGGGGGATTTGGAGTGGTG 775

Qy 201 TyrLysGlyTyrValAlaAsnAsnThrThrValAlaValLysLysLeuAlaAlaMetValAsp 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||::::
Db 776 TACAAGGGCTGTGTAGCAAACAACACCATCGTGGCAGTGAAGAAGCTGGCGGCGATGGTTGAA 835

Qy 221 IleThrThrGluGluLeuGlnGlnPheAspGlnGluIleIleLeuValAlaLysCys 240
        ||:::||||||||:|||||||||||||||||||||||||||||||||:::::||||
Db 836 ATCAGTACTGAAGAACTAAAGCAACAGTTTGATCAAGAAATTAAAGTAATGGCAAGTGT 895

Qy 241 GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspLeuCysLeu 260
        |||||||||||||||||||||||||:::|||||||||||||:::|||:::||||||
Db 896 CAGCACGAGAACCTGGTGGAGCTCCTCGGCTTCTCCAGTGACGGAGACCTGTGCTTA 955

Qy 261 ValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyThr 280
        ||||||||||||||||||||||||||||||||||:::||||||:::|||||||||||
Db 956 GTGTATGTCTACATGCCCAACGGGTCCTTGCTGGACAGACTGTCCTGCCTGGATGGTACA 1015

Qy 281 ProProLeuSerTrpHis-MetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAs 300
        |||||||||||||||||||||||||:|||||||||||||||||||||||||::
Db 1016 CCACCGCTTTCCTGGCACATGAGGTGCAAGGTTGCTCAGGGTGCAGCAAATGGCATCAG 1075

Qy 301 nPheLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAs 320
        |||||||||||||:||||||||||||||||||||:::||||||||||||||||||||||
Db 1076 GTTTCTGCATGAAAATCATCACATTCATACGAGATATTAAAAGTGCAAATATCTTACTAGA 1135

Qy 321 pGlyAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLysPheAl 340
        |:::|||||||||||||||||||||||||:::|||||||||:::||||||||||||||
Db 1136 CAAAGACTTTACTGCCAAAATATCTGACTTTGGGCTGGCTAGGGCTTCGGAAGAGTTTGCA 1195

Qy 341 aGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLe 360
        ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db 1196 GCAGACGGTCATGACCAGCCGAATCGTGGGCACCACGGCGAAGCTTT 1255

Qy 361 uArgGlyGlnIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeuGluIl 380
        |||||||||||||||||||||||||:|||||||||||||||||||||:::
Db 1256 GCGCGGGAGAAATAACCACCCAAATTGACATCTACAGCTTCGGCGTGGTTCTGTTGGAGCT 1315

Qy 381 eIleThrGlyLeuProLeuValAspAsnGlnHisAlaHisGlnProGlnLeuLeuIleAspleLy 400
        ||||||||||||||:::|||||||||||||||||||||||||||||:::
Db 1316 GATAACCGGACGCCTGGCGTGTGATAACCAACACTACTGTCGGATATTAA 1375

Qy 401 sGluGlyLeuGluAspGlyLysThrThrIleGluAspLysLysSerMetAsnAs 420
        ||||||||||||||||||:::|||||||||||||||||||||||||||:::||
Db 1376 AGAAGGAGATTGAAGATGAAGAAGACGAATTGAAGATTACACGGATGAGAGCCGA 1435

Qy 421 pAlaAspSerThrSerValGluValAlaMetTyrSerValAlaSerGlnCysLeuHisGluLy 440
        |:::|||||||||||||||||:::|||||||||||||||||||:||||||||||||||
Db 1436 TGCCGACCCTGCTTCGGTGGAAGCAAATGACTCTGCCTGCTAGCACGTGTCTGCATGAGAA 1495

Qy 441 sLysAsnLysArgProAspIleLeuLysValGlnGlnLeuLeuGlnGluMetThrAla 459
        ||||||||||:||||||||||||||||||||:||:||||||||||||::||
Db 1496 GAAAAACAGAACCGCCAGACATTGCAAAGGTTCAACAGCTCCTACAAGAGATGTCTGCT 1553

```
BG164491
LOCUS       BG164491          859 bp    mRNA    linear    EST 06-FEB-2001
DEFINITION  602342026F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4452055 5',
            mRNA sequence.
ACCESSION   BG164491
VERSION     BG164491.1  GI:12671194
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 859)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10240 row: d column: 08
            High quality sequence stop: 634.
FEATURES             Location/Qualifiers
     source          1..859
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4452055"
                     /tissue_type="hypernephroma, cell line"
                     /clone_lib="DH10B (phage-resistant)"
                     /lab_host="NIH_MGC_89"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT-primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT     266 a    168 c    190 g    235 t
ORIGIN
```

```
US-10-001-254-16 (1-460)  x  BG164491 (1-859)

Alignment Scores:
Pred. No.:
Score:                   1.23e-115
Percent Similarity:      1094.50              Length:            859
Best Local Similarity:   91.05%               Matches:           228
Query Match:             88.72%               Conservative:      6
DB:                      45.87%               Mismatches:        19
                         10                   Indels:            8
                                              Gaps:              2
```

```
Qy     1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20

Qy    21  LysLeuSerAspPheIleLeuAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    35  ATGAACAAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG  94

Qy    41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluValLeuLeuLeu  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    95  AAGCTGTCAGATTTTATTGATCCCTCAAGAGAGGATGGAAGAAGTTAGCTGTTAGCTATTAAA  154

Qy    61  GlnThrGlyLysSerProThrSerGluLeuValLeuPheAlaSerLeuLeu  80
          ||||||||||||||||||||||||||||||||||||||||||||
Db   155  AAACCATCTGGTGATGATGATACAAATCAGTTTCACATAAGGACAGATTTGAAGCATTACTT  214

Qy    81  ValGlyAlaSpLeuValAlaSpLeuLeuGlnAsnGluPhePheAlaProAlaSerLeuLeu  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   215  CAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCCACCACAAATGCACA  274

Db   275  GTTGGTTGATCTTGTGTGGATCTTTGATCCAAAATGAATTTTTTGCTCCTCTGCCAGTCTTTTG  334
```

```
RESULT 4
BF694134
LOCUS       BF694134          682 bp    mRNA    linear    EST 22-DEC-2000
DEFINITION  602082746F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4247171 5',
            mRNA sequence.
ACCESSION   BF694134
VERSION     BF694134.1  GI:11979542
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 682)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1063 row: k column: 12
            High quality sequence stop: 617.
FEATURES             Location/Qualifiers
     source          1..682
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4247171"
                     /clone_lib="DH10B (T1 phage-resistant)"
                     /lab_host="NIH_MGC_81"
                     /note="Organ: muscle (skeletal); Vector: pDNR-LIB
                     (Clontech); Site_1: SfiI (ggccgcccggcc); Site_2: SfiI
                     (ggccatcatggcc); 5' and 3' adaptors were used in cloning
                     as follows: 5' adaptor sequence:  5'-CACGCGCCATTATGGCC-3'
                     and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
```

```
Qy   101  LeuProAspAlaValProLysThrAlaAsnThrLeuPheProSerLysGluAlaIleThrVal  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   335  CTCCCAGATGCGTCGTTCCCAAAACTGGCTAATACAACTACCTTTCTAAAAGAAGCTATAACAGTT  394

Qy   121  GlnGlnLysGlnMetProPheCysAspGlyAspGlyAspArgThrLeuMetThrProValGlnAsn  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   395  CAGCAAAAACAGATGCCTTTCTGTGACAGACGATGACAGGGACATTGATGACACCTGTGCAGAAT  454

Qy   141  LeuGluGlnSerTyrMetProProAspSerSerSerProGlnAsnLysSerSerLeuGluVal  160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   455  CTTGAACAAAGCTATATGCCACCTGACTCCTCAAGTCCACAAAATAAATCTTCTAGAGTT  514

Qy   161  SerAspThrArgPheHisSerPheSerPheTyrGluLeuLeuLeuAsnValThrAsnAsnPhe  180
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   515  AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGAACAAGTGTCACAAAATAACTTT  574

Qy   181  AspGluArgProIleSerValGlyGlyAsnLysMetGlyValGlyGlyGlyGlyGlyValVal  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   575  GATGAACGACCCATTTCTGTTGGTGGTAATAAAATGGGAGGAGGGGGAGGGGTGGTAGT  632

Qy   201  TyrLysGlyTyrValAsnAsnThrThrValAlaLysValLysLeuAlaAlaMetValAsp  220
          ||||||||||||||::::||||||||||||||||||||||||||||||||
Db   633  TATAAAGGCTACGTAAATTACACAAGTGGGCAGTGAAGAAGCTTGCAGC-ATGGTGAC  691

Qy   221  IleThrThrGluGluLeuLysGlnGlnPheAspGlnIleLeuLysValMetAlaLysCys  240
          |||::::||||||||||||||||||||||||       :::     ||||||||||||
Db   692  CTTACTACTGGAGAACTGAACCAGCGTTTGTCCACGA---CTTCACGTTATGCCAG-TGT  747

Qy   241  GlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAsp  257
          |||||||                          :::||||||||||::
Db   748  TCACTTGAAAACTGTGAAACCC---------------TGGTCTCCAGGAGGGGGATAAC  789
```

```
BASE COUNT      222 a      119 c      142 g      198 t      1 others
ORIGIN
```

Alignment Scores:
Pred. No.:
Score: 2.26e-105
Percent Similarity: 1004.50
Best Local Similarity: 82.81%
Query Match: 82.81%
DB: 42.10%

```
Length:          682
Matches:         212
Conservative:    0
Mismatches:      12
Indels:          34
Gaps:            1
```

US-10-001-254-16 (1-460) x BF694134 (1-682)

```
Qy   122 GlnLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsnLeu 141
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CAAAAACAGATGCCTTTCTGTGACAAAGACAGGACATTGATGACACCCTGTGCAGAATCTT 60

Qy   142 GluGlnSerTyrMetProProAspSerSerSerSerProGluAsnLysSerLeuGluValSer 161
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GAACAAAGCTATATGCCACCTGACTCCTCCAAGTCCCAGAAAATAAAAAGTTTAGAAGTAGT 120

Qy   162 AspThrArgPheHisSerPheTyrGluLeuLysAsnValThrAsnAsnPheAsp 181
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GATACAACGTTTCACAGTTTTTCATTTATGAATTGAAGAATGTCACAAATAACTTTGAT 180

Qy   182 GluArgProIleSerValGlyGlyAsnLysMetGlyGlyGlyPheGlyValValTyr 201
         ||||||||||||||||||||||||||||||||||||||||||
Db   181 GAACGACCCAATTTCT- 195

Qy   202 LysGlyTyrValAsnAsnThrThrValAlaValLysLeuLysSerLeuLeuAla-AlaMetValAspIl 221
         ||||||||||||||||||||||||||||||||||||||||||
Db   196 ------------------GNTGGTGATGGTTGACAT 213

Qy   221 eThrThrGlyGluGluLeuLysGlnGlnIlePheGlnGluHisLysAspGlnIleLeuLysValMetAlaLysCysGl 241
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   214 TACTACTGAAGGAGAACTGAAACAGCAGTTTGATCAAGAAANTAAAAGTAAATGGCAAAGTGTCA 273

Qy   241 nHisGluAsnLeuValGlyGluGluLeuLysGlyPheSerSerAspGlyValProLeuCysLeuVa 261
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   274 ACATGAAAAACTTAGTAGGAGAACTACTGGTTTCTCAAGTGATGGAGAGACCTCGGCTTAGT 333

Qy   261 lThrValTyrMetProAsnGlySerLeuAspArgLeuSerCysLeuAspGlyThrPr 281
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   334 ATAYGTTACATGCCTAAYGGTTCATTGGTGACAGACCTCTGCTTGGATGTGTACTCC 393

Qy   281 oProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsnPh 301
         ||||||||||||||||||||||||||||||||||||||||||||
Db   394 ACCACTTTCTTGGGCACATGAGATGCAAGATTGCTCAGGGTGCAGCTAATGGCATCAATTT 453

Qy   301 eLeuHisGluAsnHisHisIleHisArgAspIleLysSerAlaAsnIleLeuLeuAspGl 321
         ||||||||||||||||||||||||||||||||||||||||
Db   454 TCTACATGAAAATCATCATANTTCATAGAGATATTAAAAGTGCAAANATATCTTACTGGATGA 513

Qy   321 uAlaPheThrAlaLysIleSerAspPhe-GlyLeu-AlaArgalaSerGluLysPheAla 340
         ||||||||||||||||||||||||||
Db   514 AGCNTTTACTGCTAAAAATNTCNGACTNTGGNCTTCGCACGGGCTTCTCGAGAAGTTTGCC 573

Qy   341 GlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeu 360
         |||||||||||||||||||||||||||||||||
Db   574 CAGACAGTGCATGACTAGCAGAATGGGAACAACAGT-TATATGGGACAGAGAGCTTTG 632

Qy   361 ArgGlyGlyIleThrProLysSerAspIleTyrSerPheGly 374
         |||||||||||||||||||||||||||||||||
Db   633 CGTGCGAGAAATTAAACCAAA-TCTGATATTAACAGGTTGGGG 673
```

Alignment Scores:
Pred. No.:
Score: 1.41e-98
Percent Similarity: 946.00
Best Local Similarity: 93.24%
Query Match: 88.99%
DB: 39.65%

```
Length:          718
Matches:         184
Conservative:    9
Mismatches:      14
Indels:          0
Gaps:            0
```

US-10-001-254-16 (1-460) x BQ780493 (1-718)

```
Qy   253 SerAspGlyAspAspLeuCysLeuValThrValTyrMetProAsnGlySerLeuAspArgLeuAsp 272
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   716 AGCGACAGTGACAACCTGTGCTTAGTGTAYGCTACATGCCCNAATGGGTTCCTTGCTTAGAC 657
```

```
Qy    273 ArgLeuSerCysLeuAspGlyThrProLeuSerTrpHisMetArgCysGlyIleAla 292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    656 AGACTGTCCTGCCTGGACGGAACGCCACTTCCTGGCACATGAGATGCAAGATTGCT 597

Qy    293 GlnGlyAlaAlaAsnGlyIleAsnPheLeuHisGluAsnHisHisIleHisArgAspIle 312
          |||||||||||||||::|||||||||||||||||||||::||||
Db    596 CGGGGTGCAGCAAACGCCATCCGTTTTTTGCATGAAAATCATCATAATTCACAGAGATATT 537

Qy    313 LysSerAlaAsnIleLeuLeuLeuAspGluValAlaPheThrAlaLysIleSerAspPheGlyLeu 332
          ||||||||||||||||::::|
Db    536 AAAAGTGCAAATAATCTTACTCAGACTTTACTGCCAAAATATCTGACTTGGGGCTT 477

Qy    333 AlaArgAlaSerGlyLysPheAlaGlnThrValMetThrSerArgIleValGlyThrThr 352
          |||||||||||||||||||||
Db    476 GCACCGGCGCTTCCGCGGAAGCTTCGCCAAAAACGGTCATGACCAGCCGAATTGTGGGCCACAACG 417

Qy    353 AlaThrMetAlaProGlyAlaLeuArgGlyValGlyIleLeuThrProLysSerAspPheTyrSer 372
          ||||||||||||||||:::||
Db    416 GCTTATATGGGCACCTGAAGGCTTTGCCAGGAAAATAACCCCAAAACTGACATCTTACAGC 357

Qy    373 PheGlyValValLeuLeuGluIleLeuIleGlyLeuProAlaValAspGluHisArgGlu 392
          ||||||||||||||||||||
Db    356 TTCGGTGTGGTTTTATTGGAGCTGATAATTGGCCTGCCTGTGATGAGAAGTAGAGAGA 297

Qy    393 ProGlnLeuLeuLeuAspPheIleLysSerGlyGluGluGluGluGluLysThrIleGluGluAsp 412
          ||||||||||||||:::||||||
Db    296 CCTCAACTACTACTGGATATTATAAAGAGAGAGATTGAAGACGGAGGAGGAAGACGATTCGAAGAT 237

Qy    413 TyrIleAspLysLysMetAsnAspAlaAspSerThrSerValGluAlaMetTyrSerVal 432
          ||||||||||:::|||||||||
Db    236 TACACAGACGAAGAAGATGAGGCCGATGCGAGCCCGTCGGTTCGGTGACTCTGTT 177

Qy    433 AlaSerGlnCysLeuHisGlyLysLysAsnGlyArgProAspIleLysLysLysValGlnGln 452
          ||||||||||||||||||||:::|||||||||
Db    176 GCTAGCCAGTGTCTCGACGAAGAAAAACAGACGCCCAGACATTGCAAAGGTTCAACACAG 117

Qy    453 LeuLeuGlnGluMetThrAla 459
          ||:::|||
Db    116 CTGCTACAAGAGACTTCTGCA 96
```

```
RESULT 6
LOCUS         AK020397                1161 bp    mRNA    linear    HTC 05-DEC-2002
DEFINITION    Mus musculus adult male diencephalon cDNA, RIKEN full-length
              enriched library, clone:9330209D03 product:interleukin-1
              receptor-associated kinase 4 [Mus musculus], full insert sequence.
ACCESSION     AK020397
VERSION       AK020397.2  GI:151919911
KEYWORDS      HTC; CAP trapper.
SOURCE        Mus musculus (house mouse)
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1
  AUTHORS     Carninci,P. and Hayashizaki,Y.
  TITLE       High-efficiency full-length cDNA cloning
  JOURNAL     Meth. Enzymol. 303, 19-44 (1999)
  PUBMED      10349636
REFERENCE     2
  AUTHORS     Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
              Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE       Normalization and subtraction of cap-trapper-selected cDNAs to
              prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL     Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE     20499374
  PUBMED      11042159
REFERENCE     3
  AUTHORS     Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
              Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
              Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
              Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
              Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
              Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
              Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE       RIKEN integrated sequence analysis (RISA) system-384-format
              sequencing pipeline with 384 multicapillary sequencer
  JOURNAL     Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE     20530913
  PUBMED      11076861
REFERENCE     4
  AUTHORS     Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
              Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
              Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
              Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
              Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
              Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
              Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
              Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
              Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
              Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
              Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
              Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
              Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
              Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
              Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
              Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
              Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C.A., Wilming,L.,
              Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
              and Hayashizaki,Y.
  TITLE       Functional annotation of a full-length mouse cDNA collection
  JOURNAL     Nature 409 (6821), 685-690 (2001)
  MEDLINE     21085660
  PUBMED      11217851
REFERENCE     5
  AUTHORS     The FANTOM Consortium and the RIKEN Genome Exploration Research
              Group Phase I & II Team.
  TITLE       Analysis of the mouse transcriptome based on functional annotation
              of 60,770 full-length cDNAs
  JOURNAL     Nature 420, 563-573 (2002)
REFERENCE     6    (bases 1 to 1161)
  AUTHORS     Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
              Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
              Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
              Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
              Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
              Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
              Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
              Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
              Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
              Suzuki,H., Tagami,M., Tagami,A., Takahashi,F., Tanaka,K.,
              Tejima,Y., Toya,T., Yamamura,Y., Yasunishi,A., Yoshida,K.,
              Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE       Direct Submission
  JOURNAL     Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
              Physical and Chemical Research (RIKEN), Laboratory for Genome
              Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
              RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
              Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
              URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
              Fax:81-45-503-9216)
COMMENT       On Oct 4, 2001 this sequence version replaced gi:12860970.
              Please visit our web site (http://genome.gsc.riken.go.jp/) for
              further details.
              cDNA library was prepared and sequenced in Mouse Genome
              Encyclopedia Project of Genome Exploration Research Group in Riken
              Genomic Sciences Center and Genome Science Laboratory in RIKEN
              Division of Experimental Animal Research in Riken contributed to
              prepare mouse tissues. First strand cDNA was primed with a primer
              [5' GAGAGAGAAGTCAAGACGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
              prepared by using trehalose thermo-activated reverse transcriptase
              and subsequently enriched for full-length by cap-trapper. cDNA went
              through two rounds of normalization to Rot = 20.0 and subtraction
              to Rot = 370.4. Second strand cDNA was prepared with the primer
              adapter of sequence [5'
              GAGAGAGAGATTCTCAGTTAATTAAATTAATTAATCCCCCCCCCCCCCCCCC 3']. cDNA was cleaved
```

with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
end: BamHI. Host: DH10B.

```
FEATURES          Location/Qualifiers
     source       1..1161
                  /organism="Mus musculus"
                  /mol_type="mRNA"
                  /strain="C57BL/6J"
                  /db_xref="FANTOM_DB:9330209D03"
                  /db_xref="MGI:1911393"
                  /db_xref="taxon:10090"
                  /clone="9330209D03"
                  /sex="male"
                  /tissue_type="diencephalon"
                  /clone_lib="RIKEN full-length enriched mouse cDNA library"
                  /dev_stage="adult"
     CDS          183..776
                  /note="unnamed protein product; interleukin-1
                  receptor-associated kinase 4 [Mus musculus]
                  putative"
                  /codon_start=1
                  /protein_id="BAB32090.2"
                  /db_xref="GI:15919912"
                  /db_xref="MGI:1924812"
                  /translation="MNKPLTTPSTYIRNINVGIIRKLSDFIDPQBGWKKLAVAIKKPSG
                  DDRYNQFHIRRFEALLQTGKSSPTCELLFDWGTTNCTVGDLVDLLVQIELFAPATLLLP
                  DAVPQTVSLLPPKEAGNSGTNTRALSGKGQDIRNAYAXTRTQLRATGLLSPDNRSVES
                  SDTRFHSFSFHELKSITNNFDEQPASAGGNRMGBGGF"
BASE COUNT    317 a    296 c    308 g    240 t
ORIGIN
```

US-10-001-254-16 (1-460) x AK020397 (1-1161)

Alignment Scores:

Pred. No.:	2.79e-98
Score:	946.00
Percent Similarity:	87.45%
Best Local Similarity:	78.24%
Query Match:	39.65%
DB:	11

Length:	1161
Matches:	187
Conservative:	22
Mismatches:	29
Indels:	2
Gaps:	0

```
Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
Db  183 ATGAACAAGCCCGTTGACACCATCGACACATACCGAACCTTAATGTGGGGAATCCTTAGG 242

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
Db  243 AAGCTGTCCGATTTTATTGATCCTCAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAA 302

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu  60
Db  303 AAGCCGTCCGGCGACGACGAATACAATCAGTTCCATATAAGGAGATTCGAAGCCTTACTT 362

Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr  80
Db  363 CAGACCGGAAAGAGCCCCACCTGTGAACTGCTGTTTGACTGGGGCACCACGAACTGCACA 422

Qy   81 ValGlyAspLeuValAspLeuLeuValGlnIleGluLeuPheAlaProAlaSerLeuLeu 100
Db  423 GTTGGCGACCTTGTGGATCTACTGGTCCAGATTGAGCTGTTTGCCCCGCACTCTCTG    482

Qy  101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGlyAla-IleThrVa 120
Db  483 CTGCCGGATGCCGTTCCCAAACCGTCAAAGCTGCTCCTAAGAAGGGGGCAACAGT     542

Qy  120 lGlnGlnIleLysGlnMetProPheCysAspLysSerAspArgThrIleuMetThrProValGlnAs 140
Db  543 GGCACAAACACACGGGCCTTGTCAGGAAGAGGACATCCGTAATGCCTATCCAA       602

Qy  140 nLeuGluGlnSerTyrMetProProAspSerSerSerProLeuAsnLysSerLeuGluVa 160
Db  603 ACTAGAACACAGCTCGAGCACCCGACTCCTC-AGCCAGACAACAGAAGTGTAGAGTC   661

Qy  160 lSerAspThrArgPheHisSerPheSerPheHisGluLeuLysSerAsnValThrAsnAsnPh 180
Db  662 CAGCGACACACTCGGTTCCACAGCTTTTCGTTCCATGAAAACTGAGAGACATCAAACACTT 721

Qy  180 eAspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyGlyPheGlyValVa 200
Db  722 CGACGAGCAACCCCGTCTCCGTGGGCAACCGGATGGGAGAGGGGGGATTTGAGTGGTGGT 781

Qy  200 lTyrLysGlyTyrValAsnAsnThrThrValAlaValAlaIleLysGlyLeuAlaAlaMetVa 220
Db  782 GTACAAGAGGGCTGTGTGAACAACACCACTGTGGCCGTGGCGATTAAAGGGATGGTTGA  841

Qy  220 lIleThrThrGluGluLysGlnLnPheAspGlnGlnIleIleLysValMetAla 238
Db  842 AATCAGTACTGAAGAACTAAAGCAACAGTTTGATCAAGAAATTAAAGTAATGGCA 896
```

RESULT 7
BQ782157/c

```
LOCUS       BQ782157          702 bp    mRNA    linear    EST 26-JUL-2002
DEFINITION  UI-R-FF0-cpj-c-20-0-UI.s1 UI-R-FF0 Rattus norvegicus cDNA clone
            UI-R-FF0-cpj-c-20-0-UI 3', mRNA sequence.
ACCESSION   BQ782157
VERSION     BQ782157.1  GI:21990629
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   MEDLINE   97044477
   PUBMED    8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Jeff Stevens
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: DISTRIBUTION: Researchers may obtain clones
            from Research Genetics (www.resgen.com).
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..702
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="UI-R-FF0-cpj-c-20-0-UI"
                     /tissue_type="Mixed tissues"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-R-FF0"
                     /note="Vector: pT7T3-Pac (Pharmacia) with a modified
                     polylinker; Site 1: EcoR I; Site 2: Not I; UI-R-FF0 is a
                     subtracted cDNA library containing the following tissue(s
                     ): Normal cartilage and SR-SWS Tumor Line. The
                     subtraction was made according to Bonaldo, Lennon and
                     Soares, Genome Research, 6.791-806, 1996. The
                     oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tags for these libraries are: CTAATGGACG,
                     CATTCTTGTA.
                     TAG_LIB-UI-R-FF0
```

BASE COUNT   158 a   169 c   156 g   219 t
ORIGIN

TAG TISSUE=cartilage
TAG_SEQ=CTAATGGACG"

US-10-001-254-16 (1-460) x BQ782157 (1-702)

Alignment Scores:
Pred. No.:
Score: 2.33e-98
Percent Similarity: 94.55%   Length: 702
Best Local Similarity: 90.58%   Matches: 183
Query Match: 13   Conservative: 8
DB: Mismatches: 11
                                        Indels: 0
                                        Gaps: 0

```
Qy    258 LeuCysLeuValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeu 277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    701 CTGTGCTTAGTGTATGCTTACATGGCTTCCTTGCCTTGAGACAGACTGTCCTGCCTG 642

Qy    278 AspGlyThrProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyValAlaAlaAsn 297
          |||||||||||||||||||||||||| |||||||:||||||||||||||||||||||||
Db    641 GACGGAACGCCACCACTTTCCTGGCACATGAGATGCAAGATTGCTCGGGGTTGCACCAAAC 582

Qy    298 GlyIleAsnPheLeuHisGluAsnHisHisHisIleHisArgAspIleLysSerAlaAsnIle 317
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    581 GGCATCCGTTTTTGCATGAAAATCATCATTATTCACGAGATATTAAAGTCAGCAAATATC 522

Qy    318 LeuLeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGlu 337
          ||||||||||| |||||||||||||||:||||||||||||||||||||||||||||||
Db    521 TTGGAGCTGATAAACCGACTTGCGCGTGTGGATGAAAACCCTGAAACCTCAACTACTACTG 462

Qy    338 AspIleLysGluGluIleLeuGluAspGluGlyLysThrIleGluAspTyrIleAspLysLys 357
          |||:|| |||||||||||||||||||||||||||||||||||||||| :||:||
Db    461 GATATTAAAGAAGAAGAGATTGAAGACGAGGACGAATTACACAGACGAGGAG 402

Qy    358 GluLeuAlaLeuArgGlyGlyIleLeuThrProLysSerAspIleTyrSerPheGlyValValLeu 377
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401 GAAGCTTTGGCAGGAGGAAATAACACCCAAATCTGACATCTACAGCTTCGGTGTGGTTTTA 342

Qy    378 LeuGluIleIleThrGlyLeuProAlaValAspAspGluHisArgGluProGlnIleLeuLeu 397
          ||||||||:|||||||||||||| |||:|||||||||||||||||||||||||
Db    341 TTGGAGCTGATAAACCGACTTGCGCGTGTGGATGAAAACCCTGAAACCTCAACTACTACTG 282

Qy    398 AspIleLysGluGluIleLeuGluAspGluGlyLysThrIleGluAspTyrIleAspLysLys 417
          |||:|||||||||||||||||||||||||||||||||||||||||:||:||
Db    281 GATATTAAAGAAGAAGAGATTGAAGACGAGGACGAATTACACAGACGAGGAG 222

Qy    418 MetAsnAspAlaAspSerThrSerValGlyLeuMetTyrSerValAlaSerGlnCysLeu 437
          |||:|||::|||||||||||||||||||||||||||||||||||||||||||||
Db    221 ATGAGCGATCGGACCCCGGCGTCGGAGGCGATGTACTCCGTTGGTCAGTGTCTG 162

Qy    438 HisGlyLysAsnLysAlaArgProAspIleLysLysValGlnGlnIleLeuGlnGluMet 457
          |||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db    161 CACGGAAGAAAAAACAGACGCGCCACATTGCAAAGGTTCAACAGCTGCTACAGAGAC 102

Qy    458 AspIleLysAsnGluIleLeuGluAspGluGlyLysThrIleGluAspTyrIleAspLysLys 461
          |||:||||
Db    461 AAGCTTGCACAAACGGTCATGCACCAGCCGAATTGTGGGCACAACGGCTTATATGCACT 402
```

LOCUS   BG616438   811 bp   mRNA   linear   EST 18-APR-2001
DEFINITION 602642772F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4773760 5',
            mRNA sequence.
ACCESSION   BG616438
VERSION   BG616438.1 GI:11667809
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 811)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.

FEATURES   Location/Qualifiers
     source   1..811
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4773760"
                     /tissue_type="embryonal carcinoma"
                     /clone_lib="NIH_MGC_61"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccattatggcc); Site_2: SfiI (ggccatatggcc).
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts and was constructed by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     library."

BASE COUNT   246 a   196 c   164 g   205 t
ORIGIN

US-10-001-254-16 (1-460) x BG616438 (1-811)

Alignment Scores:
Pred. No.:
Score: 2.75e-97
Percent Similarity: 81.10%   Length: 811
Best Local Similarity: 77.95%   Matches: 198
Query Match: 10   Conservative: 8
DB: Mismatches: 32
                                        Indels: 16
                                        Gaps: 4

```
Qy      1 MetAsnLysProIleThrProSerThrTyrValArgCysLysLeuAsnValGlyLeuIleLeuArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAACAAAACCATAACACCATCAACATAGTGCCGTGCCTGCTCAAATGTTGGACTAAATAGG 53

Qy     21 LysLeuSerAspPheIleAspProGlnGlyLysThrProLysGlyLeuAlaValAlaAlaLys 40
          |||||||||||||||:||||:|||||||||||||||||||||||||||||||||
Db     53 AAGCTGTCAGATTTTATTGATCCTCAAGGAAAGACTCCAAAGGATGGAGCAGTTAGCTGTAGCTATTAAA 113

Qy     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluValAlaLeuLeu 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGGTTTGAAGCATTACT 173

Qy     61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspProGlyThrThrAsnCysThr 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    173 CAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACCCGGGCACCACAAATTGCACA 233

Qy     81 ValGlyAspLeuValAspProLeuLeuLeuIleGlnAsnGluPheProAlaSerLeuLeu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233 GTTGGTGACTTGGTGGATCCTTTGCTCCTAATCCAAAATGAATTTCCTGCGAGTCTTTG 293

Qy    101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 CTCCCAGATGCTGTTCCCAAAACTGCTAATACACTACCTTCTAAAGAAGCTATAACAGTT 353

Qy    121 GlnGlnLysGlnMetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsn 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 CAAAACTGGTTT 412
```

```
Db   413 CAGCAAAAACAGAYGCCTTTCTGTGACAAAGACAGGACATTGATGACACCTGTGCAGGAAT 472
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   141 LeuGluGlnSerTyrMetProProAspSerSerProGluAsnLysSerLeuGluVal 160

Db   473 CTTGAACAAAGCTATATGCCACCTGACTCCTCAAGTCCAGAAAACTACAAGTTTAGAAGTT 532
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   161 SerAspThrArgPheHisSerPheSerPheTyrGlu-LeuLysAsnValThrArgAsn-P 180

Db   533 AGYGATACACGTTTTCACAGTCYTTCATTTCATCGACACAGAATGTCCACAAATAACTC 592
           :::::|||||
Qy   180 heAspGluArgProIleSerValGlyIleGlyAsnGlyGlyGlyGlyPheGlyVal 199

Db   593 TCGATGAACGAGACCCATTTCTGTCCGGGTTACTAACATGCGGAGGAGTTCGGCGAG 652
           ||||||||||||||||
Qy   200 ValTyrLysGlyTyrValAsn-------AsnThrThrValAlaValLysLysLeuAla 216

Db   653 TTGTATCATTCACGGCTAAATACACCCCAACTGTGCCCCCCGGGAAAGAAAAGT 712
           :::|||||
Qy   217 Ala-------MetValAspIleThrThrGluGluLeuGlnGlnPhe--------- 230

Db   713 TGCAGCCACCGGGTTGACATT-------CACTACCGGCCAACAACTCGTTACAGACACCT 763
           :::||||||
Qy   231 ---AspGlnGluIleLeuVsValMetAlaAlaLysCysGln 241

Db   764 TGATCCAGAGACACTAACGGCACTGGGCCCAAGTGTCAA 799
           :::|||||
```

```
RESULT 9
BU441365
LOCUS       BU441365                769 bp    mRNA    linear   EST 29-NOV-2002
DEFINITION  603208981F1 CSEQRBN11 Gallus gallus cDNA clone ChEST185a21 5', mRNA
            sequence.
ACCESSION   BU441365
VERSION     BU441365.1  GI:25930676
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
ORGANISM    Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 769)
AUTHORS     Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE       A Comprehensive Collection of Chicken cDNAs
JOURNAL     Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE     22335534
PUBMED      12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST
            )
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 0161200 8930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
            Noti adapters, digested with EcoRI, size-selected, and
            cloned into the NotI and EcoRI compatible sites of a
            custom modified MCS of the pBluescript (KS+) vector. The
            library was normalized in 2 rounds using conditions
            adapted from Soares et al., PNAS (1994) 91: 9228-9232 and
            Bonaldo et al., Genome Research 6 (1996): 791, except that
            a significantly longer reannealing hybridization was
            used."
FEATURES             Location/Qualifiers
     source          1..769
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Layer and broiler"
                     /db_xref="taxon:9031"
                     /clone="ChEST185a21"
                     /sex="Male and female"
                     /tissue_type="muscle"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQRBN11"
                     /note="Vector: pBluescript II KS(+); Site_1: EcoRI;
                     Site_2: NotI; This normalized library was constructed from
                     1 million independent clones. cDNA synthesis was initiated
                     using an oligo(dT) primer, using methylated C in the first
                     strand synthesis reaction. Following this first strand
                     reaction, double-stranded cDNA was blunted, ligated to
BASE COUNT      230 a    146 c    180 g    213 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:              3.29e-90        Length:        769
Score:                  874.00          Matches:       166
Percent Similarity:     87.78%          Conservative:   28
Best Local Similarity:  75.11%          Mismatches:     26
Query Match:            36.63%          Indels:          1
DB:                     13              Gaps:            0
```

```
US-10-001-254-16 (1-460) x BU441365 (1-769)

Qy   234 IleLysValMetAlaLysCysGln-HisGluAsnLeuValGluLeuLeuGlyPheSerSe 253
           ||||   :||||||||||||||    ::||   ||  ||||  ||::||| |||||
Db     1 CTAGAAATTATGGCAAAGTGTAAACAATGAAAATCTGGTTAGAAATTGCTCA 60

Qy   273 rLeuSerCysLeuLeuAspGlyThrProProLeuSerTrpHisMetArgCysGlyLeuAl 293
Db    61 TGATGGGTGCTCAGCCCCTGTTCGGTGTAGAACATACATGCCCAATGGTTCCGCTGCTTGACAG 120

Qy   253 rAspGlyAspAspLeuCysLeuVsValThrTyrValTyrMetProAsnGlySerLeuLeuAspPr 273
Db   121 ACTTGCTTGTCTCTGGATGGCAACTCCACCTATTTCTTGGAACACAAGGTGTGAAAATTGCTCA 180

Qy   293 nGlyAlaLeuAsnGlyIleIleAsnPheLeuAsnHisHisHisIleHisArgAspIleLeLy 313
Db   181 AGGTACTGCGAATGGCATCAACAACTTTCTCGATGACAATAATCATTATTCACAGAGACATTAA 240

Qy   313 sSerAlaAsnIleLeuLeuAspGluAlaPheThrAlaLysIleLeSerAspAspPheGlyLeuAl 333
Db   241 AAGGTGCAAATATCTTATTAACTGATGACGTATGTGCCAAAAATTCGGACTTTGGACTTGC 300

Qy   353 aTyrMetAlaProGluIleAlaLeuArgGlyValLeuThrProLysSerAspIleTyrSerPh 373
Db   301 AAGAGCCATCAGTAACATTCACACGACAACCATCAAGAGACAGGTTGTGGGAACAGCAGC 360

Qy   373 eGlyValLeuLeuLeuGluLeuIleThrGlyLeuProAlaValAspGluHisArgGluPr 393
Db   361 CTATAACGGCACCTGAGCCTCCGGAGGAGAAATAACGCCTAAAATCCGGATATCTTCAGCTT 420

Qy   393 oGlnLeuLeuLeuAspPheIleGluGluIleGluGlyLysLysThrHisIleGluAspTy 413
Db   421 TGGGGTAGTCTTACTAGAAGTAAATAAACAGGTCTCCACCAGTAGCGAAAAACCGGAGCC 480

Qy   413 rIleLeuAspLysLysMetAsnAspAlaAspSerThrSerValGluIleAlaMetTyrSerValAl 433
Db   481 ACAGTTACTGTTAAGTGGAGATGAACTTGAAGGATGAAGGAGGACTATAGAGGATTA 540

Qy   433 aSerGlnCysLeuHisGlyLysGlysAsnArgProAspIleLysLysValGlnGlnIle 453
Db   541 TGTTGACGTAAAGATGAGTGCAGCTGGGATGCAACTTCAGTTCATAAAAAATGTATTTCACTTGC 600

Qy   453 u 453
Db   601 TGATCAGTGTCTGAATGAGAACAAAAAAAACAGGAGCCCAAACATGAAGAGATGGTGTATGTAGCCT 660

Db   661 G 661
```

```
RESULT 10
BF696981
LOCUS       BF696981                719 bp    mRNA    linear   EST 22-DEC-2000
DEFINITION  602130160F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4287014 5',
```

```
ACCESSION   BF696981
VERSION     BF696981.1  GI:11982389
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1124  row: g  column: 15
            High quality sequence stop: 632.
```

mRNA sequence.

EST.

```
FEATURES             Location/Qualifiers
     source          1..719
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4287014"
                     /tissue_type="primitive neuroectoderm"
                     /clone_lib="NIH_MGC_56"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCGCCATTATGGCC-3', and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      216 a    145 c    163 g    195 t
ORIGIN
```

US-10-001-254-16 (1-460) x BF696981 (1-719)

```
Alignment Scores:
  Pred. No.:          3.6le-88
  Score:              856.00
  Percent Similarity: 91.41%         Length:           719
  Best Local Similarity: 88.10%      Matches:          185
  Query Match:        35.88%         Conservative:     7
  DB:                 10             Mismatches:       10
                                     Indels:           9
                                     Gaps:             1
```

```
Qy    1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuLeuArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 127

Qy   21 LysLeuSerAspPheIleAspProGlnGlyTrpLysLeuAlaValAlaIleLys 40
        ||||||||||||||||||||||||||||||||||||||||
Db  128 AAGCTGTCAGATTTTATTGATCCTCAAGGAGGATGGAAGTTAGCTGTAGCTATTAAA 187

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  188 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGGAGATTTGAAGCATTACTT 247

Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeu-PheAspTrpGlyThrThrAsnCysTh 80
        ||||||||||||||||||||||| |||||||||| ||||||||||||||||||||||||
Db  248 CAAACTGGAAAAAGTCCCACTTCTGAATTACTGGTTTGACTGGGGCACCACAAATTGCAC 307

Qy   80 rValGlyAspLeuValAspLeuValIleGlnAsnGluPhe-PheAlaProAlaSerLeuL 100
        ||||||||||||||||||||||||||||||||| |||||||||||||||||||||
```

```
Db  308 AGTTGGTGGATCTGGTGGATCTTTTTGATCCAAAATGAATTTGTTGGCTCCTGCAAGTCTTT 367

Qy  100 euLeuProAlaAlaValProLysThrAlaAsnThrLeuProSerLys-GluAlaLeuThr 119
        ||||||||||||||||||||||||||||||||::|||:::| |||||||||::|
Db  368 TGCTCCCAGATGCTGTTCCCAAAACTGCTAATACACTACCTTCTAAGGAAGCTATAACA 427

Qy  120 ValGlnGlnLysGlnMetProPheCysAspLysAspArgThr-LeuMetThrProValGl 139
        ||||||||||||||||||||||||||||:|::||||||||||| |||::::|||::||:
Db  428 GTTCAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGGACAGTTGATGACACCTGTGCA 487

Qy  139 n-AsnLeuGluGlnSerTyrMetProProAspSer-SerSerProGluAsnLysSerLeu 158
        | ||||||||::||||||||||:::|||||||:|| |||||||||::|||:| |||
Db  488 GAAAATCTTGAACAAAGCTATATGCCACCTGACTCTTCAAGTTCTCCAGAGAACAAATCTTTA 547

Qy  178 nAsnPheAspGluArgProIleSerValGlyGlyAsnLysMetGlyGluGlyLyPheGl 198
        : |||||||::||||||||::::|::||| ||||||| |||||::|||:||||||| ||
Db  548 GGAAGTTAGTGATACAGTGTTCACAGTTTGCATTTCATGAACTTGAAGAAGTCACGA 607

Qy  198 yValValTyrLysGly 203
        ||||||::||||||
Db  608 TACTTTGATGAAACGACCCATTTCTGTCTGGGTCGTTCATAAAAATGGGCAC-GGAGGAATTGG 666

Qy  667 G---TGATATAAAGGT 679
```

```
RESULT 11
LOCUS       BU225249
DEFINITION  603947374F1 CSEQCHN23 Gallus gallus cDNA clone ChEST902c19 5', mRNA
            sequence.
ACCESSION   BU225249
VERSION     BU225249.1  GI:25460959
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 962)
  AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
  TITLE     A Comprehensive Collection of Chicken cDNAs
  JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
   MEDLINE  22335534
   PUBMED   12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST
            )
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1..962
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST902c19"
                     /dev_stage="22"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN23"
                     /note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: NotI; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with EcoRI
                     , size-selected, and cloned into the NotI and EcoRI
```

BASE COUNT   280 a   184 c   222 g   275 t   1 others

ORIGIN

```
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996):791, except that a significantly longer
reannealing hybridization was used."
```

RESULT 12

LOCUS   BG691069   541 bp   mRNA   linear   EST 27-MAR-2003

DEFINITION   340084 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION   BG691069

VERSION   BG691069.1   GI:11932889

KEYWORDS   EST.

SOURCE   Bos taurus (cow)

ORGANISM   Bos taurus

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
```

REFERENCE   1 (bases 1 to 541)

AUTHORS   Sonstegard,T., Capuco,A.V., White,J., Van Tassell,C.P., Connor,E.E.
     , Cho,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D. and
     Quackenbush,J.

TITLE   Analysis of bovine mammary gland EST and functional annotation of
     the Bos taurus gene index

JOURNAL   Mamm. Genome 13 (7), 373-379 (2002)

MEDLINE   22135956

PUBMED   12140684

COMMENT   Contact: Sonstegard TS
```
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@lpsi.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 100 row: B column: 15
Seq primer: ATTTAGGTGACACTATAG.
```

FEATURES   Location/Qualifiers

source   1..541
```
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="BARC 5BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."
```

BASE COUNT   169 a   96 c   119 g   157 t

ORIGIN

Alignment Scores: (left column)

Pred. No.:

Score:	804.00
Percent Similarity:	68.71%
Best Local Similarity:	55.10%
Query Match:	33.70%
DB:	13

US-10-001-254-16 (1-460) x BU225249 (1-962)

Length: 962
Matches: 162
Conservative: 40
Mismatches: 37
Indels: 56
Gaps: 4

Pred. No.:   5.62e-82

```
Qy  165  PheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPheAspGluArgPro 184
              ||||||||||||||||| |||||||:||||:::||||||:||||||||||
Db    4  TTCCAAATTTTTGGTTTCACGACTTGGAAAGTGTTACAAATATAATTTTGATGAACGACCA 63

Qy  185  IleSerValGlyGlyAsnLysMetGlyGlyIleGlyGlyProGlyValValTyrLysGlyTyr 204
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db   64  GAATCAGCTGGAGGAAATAAGCTCGGGAGGAGTGGCTTTGGCATTGTTGTTCAAAGGCTAC 123

Qy  205  ValAsnAsnThrThrValAlaValAlaValLysValSerLeuAlaAlaMetValAspIleThrThrGlu 224
              ||:||||||| |||||||||||||||||||:|||||||||||||| :::
Db  124  ATCAATGGCAGAAGTTGCTGTGTCAAGAAAAACTTGCTGCTGGTTGATGTTGAGTGCACAG 183

Qy  225  GluLeuLysGlnGlnPheAspGlnIleLeuLysValValMetCysAlaLysCysHisGluAsn 244
              ::||||||||||||||||||||||||||||||||||||||||||||||
Db  184  GAATTTGAAACAGCAGTTTGATCAACGAAAATATGGCAAAGTGTAAACATGAAAAT 243

Qy  245  LeuValGlyLeuLeuGlyPheSerSerAspGlyAspGluCysSerLeuValTyrValTyr 264
              ||||||||||||||:||||||||||||||||| :::
Db  244  CTGGTAGGAATTGCTTGGTTTCTTCAGAGTGATGGTGACGAGTGTAGTGAATAC 303

Qy  265  MetProAsnGlySerLeuAsnArgLeuSerCysLeuAsnAspGlyThrProProLeuSer 284
              |||||||||||||||||||||||||||||||||||||||||||::|||
Db  264  ATGCCCAATGGGTCCCTGCTTAGACTTGCTTGTCTGGATGGCCACTCCACCTATTTCT 363

Qy  285  TrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGlyIleAsnPheLeuHisGlu 304
              |||::| | |||::|||||||||||||||::|:
Db  364  TGGAACAAGGTGTGAAATTGCTCAAGGTACTCGGAATGGCATCAAACTTTCTGCATGAC 423

Qy  305  AsnHisHisIleHisIleArgAspIleLysSerAlaAsnIleLeuLeuAspGluAlaPheThr 324
              ||::|| | :::
Db  424  AATAATCAATATTCACAGAGACCATTAAAAGC-- 453

Qy  325  AlaLysIleSerAspPheGlyLeuAlaArgAlaSerGlyLysPheAlaGlnThrValMet 344
                                                            :::
Db                                                ATCAGG 459

Qy  345  ThrSerArgIleValGlyThrThrAlaTyrMetAlaProGluAlaLeuArgGlyGluIle 364
              |||
Db  454  ---------------------------------------ACACGG 507

Qy  365  ThrProLysSerAspPheTyrSerPheGlyValValLeuLeuGluIleThrGlyLeu 384
              ||| :::
Db  460  ACACGCGTAGTGCCTGGAAGGAAATCGGTTTTCTTACCCTACGCTGCTC 507

Qy  385  ProAlaValAlaAspGluHisArgGluProGlnLeuLeuAspIleLysGluGlnIleGlu 404
              ||||| |||||||||
Db  508  ------------------------TACTCTCTAACTGCTGCAAATATC-TTATTAACT 542

Qy  405  AspGluGlyLysThrIleGluAspTyrIleGluAspLysAspAlaAspSerThr 424
              ||||||||||||| || ||||
Db  543  -------GATACGTAT 560

Qy  425  SerValGlyAlaMetTyrSerValAlaSerGlnCysGluHisGlyLysGlyAsnLysArg 444
              |||||||||||||||:::||
Db  561  GATGAGGAGGGACGACTATAGAGGATTATGTTGACGTAAAGATGAGTGACTGGGATCGAACT 620
```

Right column:

Db   621   TCAGTTCATAAAATGTATTCACTTGCTGATCAGTGTCTCAGTTGTCTGAATGAGAAAAAACAGGAGG 680

Qy   445   ProAspIleLysLys-ValGlnGlnIleLeuGlnGluMet 457
       :||:|:|| |||||||||||||| :::

Db   681   CCAAACATGAAGAAATGGGTCCAGCAGTATTTACAAGAGATA 720

Alignment Scores: (second block, right)

Pred. No.:

Score:	774.50
Percent Similarity:	88.33%
Best Local Similarity:	83.33%
Query Match:	32.46%
DB:	10

US-10-001-254-16 (1-460) x BG691069 (1-541)

Length: 541
Matches: 150
Conservative: 9
Mismatches: 19
Indels: 2
Gaps: 1

Pred. No.:   6.41e-79

```
Qy   26  IleAspProGlnGlyTrpLysLysLeuAlaValAlaIleLysGlyProSerGlyAsp 45
              ||||||||||||||||||||||||||||||||||||||||||
Db    2  ATTGATCCTCAAGAAGGATGGAAGAAGTTAGCAGTCGCTATTAAAAACCATCTGGTGAT 61

Qy   46  AspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeuGlnThrGlyLysSer 65
              ||||||||:||||||||||||||||||||||||||||||||||
Db   62  GATAGAGATACAATCAGTTTCACATAAGGAGGATTTGAAGCAATTACTGCAAATTGGAAAAGC 121
```

```
Qy   66  ProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThrValGlyAspLeuVal  85
         |||||
Db  122  CCCACGTGTGAGTTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGGTGATCTTGTG 181

Qy   86  AspLeuLeuIleGlnAsnGluPheProAlaProAlaSerLeuLeuLeuProAspAlaVal 105
         ||:|||
Db  182  GATATTTTGGTTCCAAAATGAGTTTTTGCCCCTGCCAAGTCTTTTGCTACCAGATGCTGTA 241

Qy  106  ProLysThrAlaAsnThrLeuProSerLys--GluAlaIleThrValGlnGlnLysGln 124
         ||||||
Db  242  CCCAAAAAATGTTAAATACACTCGCCTTCTAAAGTCACTTCTAGTAGCAGTTCAGCAGAAACCG 301

Qy  125  MetProPheCysAspLysAspArgThrLeuMetThrProValGlnAsnLeuGluGlnSer 144
              :::
Db  302  AAGCCCCTCTGTGGCAAAGACAGGACATCTGATATTCCTGATGAAATCCTGAACAAAAAC 361

Qy  145  TyrMetProProAspSerSerSerProGluAsnLysSerLeuGluValSerAspThrArg 164
         |:::
Db  362  TATGTGCTACCTGACCTCCTCAAGTCCAGAAAATACAAGTTTAGAAGTTTAGTGATACACGT 421

Qy  165  PheHis-SerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPheAspGluArgPr 184
         |||||  :::|||
Db  422  TTTCACAAGTTTTTCATTTTTGAATTGAAGGATGTCACAAATAACTTTGATGAACGGCC 481

Qy  184  oIleSerValGlyGlyAsnLysMetGlyGlyGlyPheGlyValValTyrLysGly 203
              ::::
Db  482  CATTTCTGTTGGTGGTAACAAAAATGGGAGAGGGGTTTGGAGTTGTGTATAAAGGC 539
```

RESULT 13
LOCUS   BB613167   676 bp   mRNA   linear   EST 26-OCT-2001
DEFINITION BB613167 RIKEN full-length enriched, 10 day neonate skin Mus
            musculus cDNA clone 4732460109 5', mRNA sequence.
ACCESSION   BB613167
VERSION   BB613167.1 GI:16453871
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 676)
AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
            Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
            M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
            D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL   Unpublished
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp;
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,K., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,
            M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,
            S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

, Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES   Location/Qualifiers
     source   1..676
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4732460109"
                     /sex="mixed"
                     /tissue_type="skin"
                     /dev_stage="10 days neonate"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 10 day neonate
                     skin"
                     /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'-
                     GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5'- GAGAGAGAGGATCCAAGAGCTCGGTTAATTAAATTAATCCCCCCC
                     CCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I"
BASE COUNT   180 a   196 c   172 g   128 t
ORIGIN

US-10-001-254-16 (1-460) x BB613167 (1-676)

Alignment Scores:
Pred. No.:   1
Score: 1.58e-76   Length: 676
Score: 755.00   Matches: 140
Percent Similarity: 88.08%   Conservative: 22
Percent Identity: 76.09%   Mismatches: 22
Best Local Similarity: 76.98%   Indels: 0
Query Match: 31.64%   Gaps: 0
DB: 10

```
Qy    1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
         ||||||||||||
Db  116  ATGAACAAGCCGTTGACACCATCGACATACATACGGTGTTTAATGTGGGAATCCTTAGG 175

Qy   21  LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaAlaLeuLys  40
         ||||:||
Db  176  AAGCTGTCGGATTTATTGATCCTCAAGAAGGGGTGGAAGAAATTAGCAGTAGCTATCAAA 235

Qy   41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu  60
         ||||:|||
Db  236  AAGCCGTCCGGCGACGACAGATACAATCAGTTCCATATAAGGAGAGATTCGAAGCCTTACTT 295

Qy   61  GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr  80
         |||||||||
Db  296  CAGACCGGGAAGAGCCCCACCTGTGAACTGCTGTTTGACTGGGGCACCACGAACTGCACA 355

Qy   81  ValGlyAspLeuValAspLeuLeuIleGlnAsnGluPheProAlaProAlaSerLeuLeu 100
         |||||||||||:||||||
Db  ...
```

US-10-001-254-16 (1-460) x BM431425 (1-478)

```
Alignment Scores:
Pred. No.:
Score:                          2.15e-76
                                752.00
Percent Similarity:             98.70%
Percent Identity:               95.45%
Best Local Similarity:          31.52%
Query Match:                    12

                    Length:        478
                    Matches:       147
                    Conservative:  5
                    Mismatches:    2
                    Indels:        1
                    Gaps:          0
```

```
BASE COUNT       139 a      94 c     106 g     139 t
ORIGIN
```

RESULT 14
BM431425

LOCUS   BM431425   478 bp   mRNA   linear   EST 31-JAN-2002
DEFINITION   1Duo16D05 Bos taurus Duodenum #1 library Bos taurus cDNA, mRNA
             sequence.
ACCESSION   BM431425
VERSION   BM431425.1 GI:18453147
KEYWORDS   EST.
SOURCE   Bos taurus (cow)
  ORGANISM   Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
             Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 478)
  AUTHORS   Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Sensen,C.W., Gordon
             ,P.M.K. and Moore,S.S
  TITLE   Gene Expression Profiling of the Bovine Gastrointestinal Tract
  JOURNAL   Unpublished
COMMENT   Contact: Dr. Stephen Moore
             Beef Genomics Laboratory
             Dept. of AFNS, University of Alberta
             410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
             Tel: 780 492 0169
             Fax: 780 492 4265
             Email: stephen.moore@ualberta.ca
             Insert Length: 478 Std Error: 0.00
             POLYA=No.
FEATURES   Location/Qualifiers
     source   1..478
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="Smooth muscle"
                     /cell_type="simple columnar epithelial"
                     /dev_stage="Young adult"
                     /lab_host="XL1-BlueMRF'strain"
                     /clone_lib="Bos taurus Duodenum #1 library"
                     /note="Organ: Intestine/duodenum; Vector: Uni-ZAPXR;
                     Site_1: EcoRI; Site_2: Xho I"
```

```
Qy    356 GTTGGGACCTTGTTGGATCTACTGGTTCAGATTGAGCTGTTTGCCCCCACTTCCTG 415
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    416 CTGCCGGATGCCGTTCCCAAACCGTCAAAAGCCTGCCTCCTTCCAGAGACGGGCAACAGTG 475

Qy    101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal 120
          |||||||||||||||||||||||||||||||||||||||||
Db    416 CTGCCGGATGCCGTTCCCAAACCGTCAAAAGCCTGCCTCCTTCCAGAGACGGGCAACAGTG 475

Qy    121 GlnGlnLysGlnMetProPheCysAspLysAspAspArgThrLeuMetThrProValAsn 140
          |||||||||||||||   ||| ||
Db    476 GCACAAACACACGGGCCCTTGTCAGGAAAAGGACAGGACATCCGTAATGCCTATCCGAAG 535

Qy    141 LeuGluGlnSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal 160
          |||||||||||||||||||||||||||||||||||||||
Db    536 CTAGAACAACAGCTGCGAGCCACCGGACTCCTCAAGCCCAGACAACAGAAGTGTAAAGTCC 595

Qy    161 SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe 180
          ||||||||||||||||||||||||||||||||||
Db    596 AGGGACACATCCGGTTCCACAACTTCTCGTTCCATGAACTGAAAAGCCATCACACAAACTTC 655

Qy    181 AspGluArgPro 184
          |||||:::|||||
Db    656 GACCAGCAACCC 667
```

RESULT 15
BF687921

LOCUS   BF687921   858 bp   mRNA   linear   EST 22-DEC-2000
DEFINITION   602069996F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4066054 5',
             mRNA sequence.
ACCESSION   BF687921
VERSION   BF687921.1 GI:11973329
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 858)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM902 row: p column: 23
             High quality sequence stop: 564.
FEATURES   Location/Qualifiers
     source   1..858
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4066054"
                     /tissue_type="globlastoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_57"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     Sfil (ggccgcctcggcc); Site_2: Sfil (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
```

```
Qy    239 LysCysGlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspAspLeu 258
          :::||||||||||||||||||||||||||||||||||||||||||
Db      2 AGGGTCAACATGAAAAACTTAGTAGAATTACTTGGTTTCAACAGTGATGGAGATGACTTC 61

Qy    259 CysLeuValThrValThrTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAsp 278
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TGCTTAGTTACTGTTACTTACATGCCCAATGGTTCAATTGCTTGACACGACTGCTTGGTTGGAT 121

Qy    279 GlyThrProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGly 298
          |||||||||||||||:::|||||||||||||||||||||||||||||
Db    122 GGTACTCCCCCACTTCTTTGGAACATGAGATGCAAGATTGCCCAGGGTGCAGCTAATGGC 181

Qy    299 IleAsnPheLeuHisGluAsnHisIleHisIleLysHisArgAspIleLysSerAlaAsnIleLeu 318
          ::::|||||||||||||||||||||||||||||||||||||
Db    182 CTCAGTTATTTACATGAAAACCATCATATTCATATAAAAAAGTGCAAAATATCTTA 241

Qy    319 LeuAspGluAlaPheThrAlaLysIleSerAspPheGlyLeuAlaArgAlaSerGluLys 338
          ||||||||||||||||||||||||||||||||||||||||
Db    242 TTAGATGAAGACTTTACAGCCAAAATAATCTGACTTTGGCCTTGCACGGGCGTTCTGAGAAG 301

Qy    339 PheAlaGlnThrValMetThrSerArgIleValGlyThrThrAlaTyrMetAlaProGlu 358
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TTTGCCCAGACAGTTCATGACTAGCCAGAATTGTGGGAACAACAGCTTATATGGCACCTGAA 361

Qy    359 AlaLeuArgGlyGluIleThrProLysSerAspIleTyrSerPheGlyValValLeuLeu 378
          ||||||||||||||||||||||||||||||||||||||||
Db    362 GCTTTGCGAGGAGAAAATTACACCCAAATCTGACATCTACAGCTTTGGTGTGGTTTTGCTA 421

Qy    379 GluIleIleThrGlyLeuProAlaValAspHisArgGlu 392
          |||||||||||||||||||||||||||||||
Db    422 GAAATA-ATAACTGGACTTCCAGCTGTGGAGATGAACACCGTGA 462
```

BASE COUNT   289 a   142 c   236 g   191 t

ORIGIN

adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
(where B = A, C, G or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 12/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

US-10-001-254-16 (1-460) x BF687921 (1-858)

Alignment Scores:
Pred. No.:
Score: 749.50
Percent Similarity: 65.67%
Best Local Similarity: 63.00%
Query Match: 31.41%
DB: 10

Length: 858
Matches: 189
Conservative: 8
Mismatches: 48
Indels: 60
Gaps: 5

Score: 9.57e-76

```
Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
        ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   37 ATGAACAAACCCATAACCACCATCAACATATAGTGCGGCTGCTC-AATGTTGGACTAATTAGG  95

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysValSerLeuAlaValAlaLeuLys  40
        |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   96 AAGCTGTCAGATTTTAT-GATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA  154

Qy   41 LysProSerGlyAspAspAlaArgTyrAsnGlnPheHisIleArgArgPheGluAlaLeuLeu  60
        |||||||||||||||||||||||||||||||||||||||||||||
Db  155 AAACCATCTGGTGATGATGCTAGATACAATCAGTTTCACATAAGA---  196

Qy   61 GlnThrGlyLysSerProThrSerGluLeuLeuPheAspTrpGlyThrThrAsnCysThr  80
                                          ::: ||||||
Db  197 ------------------TGCTGTTCCAAAACTGT---  214

Qy   81 ValGlyAspLeuValAspLeuLeuGlnAsnGluPheAlaProAlaSerLeuLeu  100
Db  214 --------------------  214

Qy  101 LeuProAspAlaValProLysThrAlaAsnThrLeuProSerLysGluAlaIleThrVal  120
                 ||||||||||||||||||||||||||||||||||||||||
Db  215 ---------AATACACTACCTTCTAAAGAAGCTATAACAGTT  247

Qy  121 GlnGlnLysGlnMetProPheCysTyrAspSerAspArgThrLeuMetThrProValGlnAsn  140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  248 CAGCAAAAACAGATGCCTTTCTGTGACAAAGACAGCAATGATGACACCTGTGCAGAAT  307

Qy  141 LeuGluGluSerTyrMetProProAspSerSerSerProGluAsnLysSerLeuGluVal  160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  308 CTTGAACAAAGCTATATGCCACCTGACTCCTCAAGTCCAGAAAATAAAAGTTTAGAAGTT  367

Qy  161 SerAspThrArgPheHisSerPheSerPheTyrGluLeuLysAsnValThrAsnAsnPhe  180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  368 AGTGATACACGTTTTCACAGTTTTTCATTTTATGAATTGAAGAAAGTTCACAAAATAACTTG  427

Qy  181 AspGluArgProIleSerValGlyGlyAsnLysMetGlyValGlyPheGlyValVal  200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  428 GATGGAAGACCCATTTCTGTCGGTGGTAATAAAATGGGAGGAGGATT-GGAGTGGGTA  486

Qy  201 TyrLysGlyTyrValAsnAsnThrThrValAlaLysAlaAlaMetValAsp  220
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  487 TATAAAGGGTACGTAAAATAACACAAT-GTGGGAGTGAAGACAGTTGAGCAATGGTTGAC  545

Qy  221 Ile-ThrThr-GluGluLeuLysGlnGlnPheAspGlnIleLysValMetAlaLysC  240
        ||| |||| |||||||||||||||||||||||||||||||||||| |||||
Db  546 ATATACTACTCGGAAGAACTGAAACAGCAGGTTGGATCAAGAAATAAG---TATGGGAAGT  602

Qy  240 ysGlnHisGluAsnLeuValGluLeuLeuGlyPheSerSerAspGlyAspLeuCysL  260
        ||| |||||||||||||||||||||||||||||||||||||
Db  603 GGGTACATGGAGAACTTAGCAGAAAATACT-GGTTCTCCAGGTGAGGGGGATAAGCCCTGGG  661

Qy  260 euValTyrValTyrMetProAsnGlySerLeuLeuAspArgLeuSerCysLeuAspGlyT  280
        ||    ::: |||
Db  662 TAGGATATGTTACAGGCCGGCAAAGGGC--------ATGGGAAACAAAACTTTTGGTGTGGGGG  715

Qy  280 hrProProLeuSerTrpHisMetArgCysLysIleAlaGlnGlyAlaAlaAsnGly  298
        ||||||||||| ||||||  ::::::||| ||||:||||:: |||
Db  716 CTCCCCCACTTTTTTGGGCATGGGAGGCCGGAATTGGGCC---GGGGGCGGCCAAGGGGGG  768
```

Search completed: January 18, 2004, 06:06:50
Job time : 4501.9 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 01:04:18 ; Search time 998.41 Seconds
                                           (without alignments)
                                           2417.512 Million cell updates/sec

Title: US-10-001-254-26
Perfect score: 323
Sequence: 1 MNKPITPSTYVRCLNVGLIR........KKPSGDDRYNQFHIRCCSQN 59

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop 6.0 , Fgapext 7.0
               Delop   6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Total number of hits satisfying chosen parameters: 5777422

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152423_19714/app_query.fasta_1.1109
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US10001254 @CGN_1_1_5671 @runat_16012004_152423_19714 -NCPU=6 -ICPU=3
 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   GenEmbl]:*

1: gb_ba:*	29: em_vi:*
2: gb_htg:*	30: em_htg_hum:*
3: gb_in:*	31: em_htg_inv:*
4: gb_om:*	32: em_htg_other:*
5: gb_ov:*	33: em_htg_mus:*
6: gb_pat:*	34: em_htg_pln:*
7: gb_ph:*	35: em_htg_rod:*
8: gb_pl:*	36: em_htg_mam:*
9: gb_pr:*	37: em_htg_vrt:*
10: gb_ro:*	38: em_sy:*
11: gb_sts:*	39: em_hgo_hum:*
12: gb_sy:*	40: em_hgo_mus:*
13: gb_un:*	41: em_hgo_other:*
14: gb_vi:*	
15: em_ba:*	
16: em_fun:*	
17: em_hum:*	
18: em_in:*	
19: em_mu:*	
20: em_om:*	
21: em_or:*	
22: em_ov:*	
23: em_pat:*	
24: em_ph:*	
25: em_pl:*	
26: em_ro:*	
27: em_sts:*	
28: em_un:*	

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	323	100.0	211	6	AX431316	AX431316 Sequence
2	323	100.0	2213	9	BD155790	BD155790 Primer fo
3	323	100.0	2213	9	AK027301	AK027301 Homo sapi
4	290	89.8	833	6	AR223870	AR223870 Sequence
5	290	89.8	1383	6	AX196260	AX196260 Sequence
6	290	89.8	1383	6	AX431306	AX431306 Sequence
7	290	89.8	1383	6	AF445802	AF445802 Homo sapi
8	290	89.8	1629	9	BC013316	BC013316 Homo sapi
9	290	89.8	2817	9	AX431318	AX431318 Sequence
10	290	89.8	2817	9	AF155118	AF155118 Homo sapi
11	290	89.8	2820	9	AK000528	AK000528 Homo sapi
12	290	89.8	33033	9	AY186092	AY186092 Homo sapi
13	290	89.8	118572	9	AC093012	AC093012 Homo sapi
14	286	88.5	165868	2	AC021719	AC021719 Homo sapi
15	277	85.8	501	6	AX321132	AX321132 Sequence
16	271	83.9	1542	6	AX196262	AX196262 Sequence
17	271	83.9	1542	10	AF445803	AF445803 Mus muscu
18	271	83.9	2431	10	BC051676	BC051676 Mus muscu
19	269	83.3	264657	5	AC129390	AC129390 Rattus no
20	247	76.5	294	5	AX431296	AX431296 Sequence
21	125	38.7	1719	5	BC045381	BC045381 Danio rer
22	77.5	24.0	190968	2	AC128035	AC128035 Rattus no
23	77.5	24.0	245669	2	AC127919	AC127919 Rattus no
24	77.5	24.0	270729	2	AC133702	AC133702 Rattus no
25	74	22.9	4320	9	AL954691_1	HSCKPSEU X64692 H.sapiens c
26	73.5	22.8	110000	10	AL683822	Continuation (2 of
27	73	22.6	187380	9	AC023946	AC023946 Homo sapi
28	71.5	22.1	2190	1	AY013286	AY013286 Bacillus
29	69.5	21.5	146812	2	AC027204	AC027204 Homo sapi
30	69.5	21.5	150855	2	AC013620	AC013620 Homo sapi
31	69.5	21.5	172914	9	CNS07ECZ	AL445383 Human chr
32	69.5	21.5	172257	9	AC137121	AC137121 Human chr
33	69	21.4	197991	5	AL929109	AL929109 Zebrafish
34	69	21.4	212293	2	AC125252	AC125252 Mus muscu
35	69	21.4	1835	8	AF500443	AF500443 Mespilus
36	68.5	21.2	5187	9	AB067497	AB067497 Homo sapi
37	68	21.1	15521	1	AB014781	AB014781 Homo sapi
38	68	21.1	122435	5	AL355481	AL355481 Homo sapi
39	68	21.1	140628	2	AC016468	AC016468 Human DNA
40	68	21.1	172280	9	AL390879	AL390879 Human DNA
41	68	21.1	175465	2	AL356783	AL356783 Homo sapi
42	68	21.1	349980	6	AX492787	AX492787 Homo sapi
43	68	21.1	349980	6	AX553954	AX553954 Sequence
44	67.5	20.9	177777	2	AC022993	AC022993 Homo sapi

ALIGNMENTS

RESULT 1

```
LOCUS       AX431316        211 bp    DNA    linear    PAT 28-JUN-2002
DEFINITION  Sequence 25 from Patent WO240680.
ACCESSION   AX431316
VERSION     AX431316.1  GI:21656185
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
            Roth,W. and Stenner-Liewen,F.
  TITLE     Novel death domain proteins
  JOURNAL   Patent: WO 0240680-A 25 23-MAY-2002;
            BURNHAM INST (US)
FEATURES             Location/Qualifiers
     source          1..211
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             1..180
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD37284.1"
                     /db_xref="GI:21656186"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFIPDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRCCSQN"
BASE COUNT       77 a      41 c      36 g      57 t
ORIGIN
```

US-10-001-254-26 (1-59) x AX431316 (1-211)

```
  Alignment Scores:
  Pred. No.:
  Score:                    4.58e-36
                            323.00          Length:       211
  Percent Similarity:       100.00%         Matches:       59
  Best Local Similarity:    100.00%         Conservative:   0
  Query Match:              6               Mismatches:     0
                                            Indels:         0
                                            Gaps:           0
```

```
Qy    1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG  60

Qy   21  LysLeuSerAspPheIleAspAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA  120

Qy   41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysCysSerGlnAsn  59
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGATGCTGTTGTTCCAAAAC  177
```

FEATURES   Location/Qualifiers
     source   1..(1864).
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"

```
FT   28-JUL-2000 JP 2000280990
PI   TOSHIO OTA, TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI   SAITO,
PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI   KEIICHI NAGAI,TETSUJI OTSUKI
PC   C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/
PC   10,
PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
     Primer for synthesizing full-length cDNA and use thereof FH Key
     CDS             (977)..(1864).
                     Location/Qualifiers
```

US-10-001-254-26 (1-59) x BD155790 (1-2213)

RESULT 2
```
LOCUS       BD155790        2213 bp    DNA    linear    PAT 17-JAN-2003
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD155790
VERSION     BD155790.1  GI:27861548
KEYWORDS    JP 2002191363-A/10633.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2213)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primer for synthesizing full-length cDNA and use thereof.
  JOURNAL   Patent: JP 2002191363-A 10633 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
  OS   Homo sapiens (human)
  PN   JP 2002191363-A/10633
  PD   09-JUL-2002
```

```
  Alignment Scores:
  Pred. No.:
  Score:                    5.1e-35
                            323.00          Length:       2213
  Percent Similarity:       100.00%         Matches:        59
  Best Local Similarity:    100.00%         Conservative:    0
  Query Match:              6               Mismatches:      0
                                            Indels:          0
                                            Gaps:            0
```

```
Qy    1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  751  ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG  810

Qy   21  LysLeuSerAspPheIleAspAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  811  AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA  870

Qy   41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysCysSerGlnAsn  59
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  871  AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGATGCTGTTGTTCCAAAAC  927
```

US-10-001-254-26 (1-59) x AK027301 (1-2213)

RESULT 3
```
LOCUS       AK027301        2213 bp    mRNA    linear    PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ14395 fis, clone HEMBA1003250, weakly similar
            to PROTEIN KINASE APKlA (EC 2.7.1.-).
ACCESSION   AK027301
VERSION     AK027301.1  GI:14041890
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2213)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
  COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology; cDNA library construction,
```

```
FEATURES
          source          Location/Qualifiers
                          1..2213
                          /organism="Homo sapiens"
                          /mol_type="mRNA"
                          /db_xref="taxon:9606"
                          /clone="HEMBA1003250"
                          /clone_lib="HEMBA1"
                          /tissue_type="whole embryo, mainly head"
                          /dev_stage="embryo, 10 weeks"
                          /note="cloning vector: pME18SFl3"
BASE COUNT      642 a         488 c         516 g         567 t
ORIGIN

          5'- & 3'-end one pass sequencing and clone selection; Helix
          Research Institute (supported by Japan Key Technology Center etc.)
          and Department of Virology, Institute of Medical Science,
          University of Tokyo.
```

US-10-001-254-26 (1-59) x AK027301 (1-2213)

```
Alignment Scores:
Pred. No.:                                      .
Score:              5.1e-35       Length:       2213
                    323.00        Matches:      59
Percent Similarity: 100.00%       Conservative: 0
Best Local Similarity: 100.00%    Mismatches:   0
Query Match:        9             Indels:       0
DB:                               Gaps:         0
```

```
Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyIleuIleArg 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   751 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 810

Qy    21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   811 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 870

Qy    41 LysProSerGlyAspAspPheHisIleArgCysCysSerGlnAsn 59
           |||||||||||||||||||||||||||||||||||||||||
Db   871 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGATGCGTTCCCAAAAC 927
```

```
RESULT 4
LOCUS       AR223870             833 bp    DNA       linear  PAT 26-SEP-2002
DEFINITION  Sequence 10 from patent US 6440663.
ACCESSION   AR223870
VERSION     AR223870.1  GI:23324452
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 833)
  AUTHORS   Scanlan,M.J., Stockert,E., Chen,Y.-T., Old,L.J., Jager,E. and
            Knuth,A.
  TITLE     Renal cancer associated antigens and uses therefor
  JOURNAL   Patent: US 6440663-A 10 27-AUG-2002;
FEATURES
          source          Location/Qualifiers
                          1..833
                          /organism="unknown"
BASE COUNT      273 a         155 c         179 g         226 t
ORIGIN
```

US-10-001-254-26 (1-59) x AR223870 (1-833)

```
Alignment Scores:
Pred. No.:
Score:              7.64e-31      Length:       833
                    290.00        Matches:      54
Percent Similarity: 100.00%       Conservative: 0
Best Local Similarity: 100.00%    Mismatches:   0
Query Match:        6             Indels:       0
DB:                               Gaps:         0
```

```
Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyIleuIleArg 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    50 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 109
```

```
Qy    21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   110 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 169

Qy    41 LysProSerGlyAspAspPheHisIleArgCys 54
           |||||||||||||||||||||||||||||||
Db   170 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 211
```

```
RESULT 5
LOCUS       AX196260             1383 bp   DNA       linear  PAT 28-AUG-2001
DEFINITION  Sequence 2 from Patent WO0151641.
ACCESSION   AX196260
VERSION     AX196260.1  GI:15386462
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Weeche.H. and Li.S.
  TITLE     Irak-4: compositions and methods of use
  JOURNAL   Patent: WO 0151641-A 2 19-JUL-2001;
            Tularik Inc. (US)
FEATURES
          source          Location/Qualifiers
                          1..1383
                          /organism="Homo sapiens"
                          /mol_type="genomic DNA"
                          /db_xref="taxon:9606"
                          /note="human IL-1 receptor-associated kinase 4 (IRAK-4)
                          cDNA"
          CDS             1..1383
                          /note="human IRAK-4"
                          /codon_start=1
                          /protein_id="CAC60090.1"
                          /db_xref="GI:15386463"
                          /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKPSG
                          DDRVNQFHIRFEALLQTGKSPTSELLFDWGTTNCTAGDLVDLLIQNEFPASLLLP
                          DAVPKTANTLPSKEAITVQQKOMPFCDKDRTLMTPVQNLEQSYMPPDSSSPENKSLEV
                          SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGVVNNTTVAVKKLAAM
                          VDITTEELKQOFDQEIKVMAKCQHENLVELLGFSSDGGDLCLVVYMPNGSLLDRLSC
                          LDGTPPLSWHMRCKIAQGAAANGINFLHNHNHIRDIKSANILLDEAFTAKISDFGLAR
                          ASEKPAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVLLEITGLPAVDEHRE
                          PQILLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                          QQLLQEMTAS"
BASE COUNT      463 a         244 c         283 g         393 t
ORIGIN
```

US-10-001-254-26 (1-59) x AX196260 (1-1383)

```
Alignment Scores:
Pred. No.:
Score:              1.28e-30      Length:       1383
                    290.00        Matches:      54
Percent Similarity: 100.00%       Conservative: 0
Best Local Similarity: 100.00%    Mismatches:   0
Query Match:        6             Indels:       0
DB:                               Gaps:         0
```

```
Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyIleuIleArg 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 60

Qy    21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 120

Qy    41 LysProSerGlyAspAspPheHisIleArg 54
           ||||||||||||||||||||||||||||||
Db   121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 162
```

```
RESULT 6
AX431306
```

Left column

```
LOCUS       AX431306                1383 bp    DNA     linear   PAT 28-JUN-2002
DEFINITION  Sequence 15 from Patent WO0240680.
ACCESSION   AX431306
VERSION     AX431306.1  GI:21656175
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
            Roth,W. and Stenner-Liewen,F.
  TITLE     Novel death domain proteins
  JOURNAL   Patent: WO 0240680-A 15 23-MAY-2002;
            BURNHAM INST (US)
FEATURES             Location/Qualifiers
     source          1..1383
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             1..1383
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD37279.1"
                     /db_xref="GI:21656176"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRREFALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQOKOMPFCDKDRTLMTPVONLEQSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTEELKQQFDQEIKVMAKCQHENLVELLGFSSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                     QOLLQEMTAS"
BASE COUNT      463 a    243 c    283 g    394 t
ORIGIN
```

US-10-001-254-26 (1-59) x AX431306 (1-1383)

Alignment Scores:

Pred. No.:	1.28e-30	Length:	1383
Score:	290.00	Matches:	54
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	89.78%	Indels:	0
DB:	6	Gaps:	0

```
Qy    1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAACAAACCCATAACACCATCAACATATGTCCGCTGCCTCAATGTTGGACTAATTAGG  60

Qy   21  LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  AAGCTGTCAGATTTTATTGATCCTCAAGAGGGATGGAAGAAGTTAGCTGTAGCTATTAAA 120

Qy   41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg  54
         ||||||||||||||||||||||||||||||||||||||||||
Db  121  AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAGG 162
```

RESULT 7

```
LOCUS       AF445802                1383 bp    mRNA    linear   PRI 20-APR-2002
DEFINITION  Homo sapiens interleukin-1 receptor associated kinase 4 (IRAK4)
            mRNA, complete cds.
ACCESSION   AF445802
VERSION     AF445802.1  GI:20219009
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1383)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
```

Right column

```
  TITLE     IRAK-4: a novel member of the IRAK family with the properties of an
            IRAK-kinase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (8), 5567-5572 (2002)
   MEDLINE  21957277
   PUBMED   11960013
REFERENCE   2  (bases 1 to 1383)
  AUTHORS   Suzuki,N., Suzuki,S., Duncan,G.S., Millar,D.G., Wada,T.,
            Mirtsos,C., Takada,H., Wakeham,A., Itie,A., Li,S., Penninger,J.M.,
            Wesche,H., Ohashi,P.S., Mak,T.W. and Yeh,W.C.
  TITLE     Severe impairment of interleukin-1 and Toll-like receptor
            signaling in mice lacking IRAK-4
  JOURNAL   Nature 416 (6882), 750-756 (2002)
   PUBMED   11923871
REFERENCE   3  (bases 1 to 1383)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) Biology I, Tularik Inc., 2 Corporate Drive,
            South San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1..1383
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..1383
                     /gene="IRAK4"
     CDS             1..1383
                     /gene="IRAK4"
                     /note="protein kinase"
                     /codon_start=1
                     /product="interleukin-1 receptor associated kinase 4"
                     /protein_id="AAM15772.1"
                     /db_xref="GI:20219010"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRREFALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQOKOMPFCDKDRTLMTPVONLEQSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTEELKQQFDQEIKVMAKCQHENLVELLGFSSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                     QOLLQEMTAS"
BASE COUNT      463 a    244 c    283 g    393 t
ORIGIN
```

US-10-001-254-26 (1-59) x AF445802 (1-1383)

Alignment Scores:

Pred. No.:	1.28e-30	Length:	1383
Score:	290.00	Matches:	54
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	89.78%	Indels:	0
DB:	6	Gaps:	0

```
Qy    1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGAACAAACCCATAACACCATCAACATATGTCCGCTGCCTCAATGTTGGACTAATTAGG  60

Qy   21  LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  AAGCTGTCAGATTTTATTGATCCTCAAGAGGGATGGAAGAAGTTAGCTGTAGCTATTAAA 120

Qy   41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg  54
         ||||||||||||||||||||||||||||||||||||||||||
Db  121  AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAGG 162
```

RESULT 8

```
LOCUS       BC013316                1629 bp    mRNA    linear   PRI 04-SEP-2001
DEFINITION  Homo sapiens, clone MGC:13330 IMAGE:4287014, mRNA, complete cds.
ACCESSION   BC013316
VERSION     BC013316.1  GI:15426431
```

```
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1629)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
COMMENT     REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 19 Row: n Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7705840.
FEATURES             Location/Qualifiers
     source          1..1629
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:13330 IMAGE:4287014"
                     /tissue_type="Brain, primitive neuroectodermal"
                     /clone_lib="NIH_MGC_56"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             71..1453
                     /codon_start=1
                     /product="Unknown (protein for MGC:13330)"
                     /protein_id="AAH13316.1"
                     /db_xref="GI:15426432"
                     /translation="MNKPITPSTVYRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPFSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANVLPSKEAITVQQKOMPFCDKDRTLWTPVVQNLEQSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTEELKQOFDQEIKVMAKCQHENLVELLGFSSSDGGDDLCLVVVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAOGRAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTPAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     POLLLDIKEEIEDEEKTIEDYIDRKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                     QOLLQEMTAS"
BASE COUNT      542 a    291 c    328 g    468 t
ORIGIN
```

US-10-001-254-26 (1-59) x BC013316 (1-1629)

```
Alignment Scores:
 Pred. No.:
 Score:                 1.52e-30     Length:         1629
 Percent Similarity:    290.00       Matches:        54
 Best Local Similarity: 100.00%      Conservative:   0
 Query Match:           89.78%       Mismatches:     0
 DB:                    9            Indels:         0
                                     Gaps:           0

Qy   1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  71  ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 130

Qy  21  LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 131  AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 190
```

```
Qy  41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg  54
        ||||||||||||||||||||||||||||||||||||||||||
Db 191  AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 232
```

RESULT 9
AX431318

```
LOCUS       AX431318                2817 bp    DNA     linear   PAT 28-JUN-2002
DEFINITION  Sequence 27 from Patent WO0240680.
ACCESSION   AX431318
VERSION     AX431318.1  GI:21656187
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Pawlowski,K., Florentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
            Roth,W. and Stenner-Liewen,F.
  TITLE     Novel death domain proteins
  JOURNAL   Patent: WO 0240680-A 27 23-MAY-2002;
            BURNHAM INST (US)
FEATURES             Location/Qualifiers
     source          1..2817
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             50..1432
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD37285.1"
                     /db_xref="GI:21656188"
                     /translation="MNKPITPSTVYRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPFSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANVLPSKEAITVQQKOMPFCDKDRTLWTPVVQNLEQSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTEELKQOFDQEIKVMAKCQHENLVELLGFSSSDGGDDLCLVVVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAOGRAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTPAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     POLLLDIKEEIEDEEKTIEDYIDRKMNDADSTSVEAMYSVASQCRHEKKNKSPDIKKV
                     HQLLQEMTAS"
BASE COUNT      912 a    547 c    586 g    772 t
ORIGIN
```

US-10-001-254-26 (1-59) x AX431318 (1-2817)

```
Alignment Scores:
 Pred. No.:
 Score:                 2.66e-30     Length:         2817
 Percent Similarity:    290.00       Matches:        54
 Best Local Similarity: 100.00%      Conservative:   0
 Query Match:           89.78%       Mismatches:     0
 DB:                    6            Indels:         0
                                     Gaps:           0

Qy   1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50  ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 109

Qy  21  LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110  AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 169

Qy  41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg  54
        ||||||||||||||||||||||||||||||||||||||||||
Db 170  AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 211
```

RESULT 10
AF155118

```
LOCUS       AF155118                2817 bp    mRNA    linear   PRI 05-JAN-2000
DEFINITION  Homo sapiens putative protein kinase NY-REN-64 antigen mRNA,
            complete cds.
ACCESSION   AF155118
VERSION     AF155118.1  GI:5360130
```

Left column

KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2817)
  AUTHORS   Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
             Jongeneel,V., Gure,A.O., Jager,D., Jager,E., Knuth,A., Chen,Y.-T.,
             and Old,L.J.
  TITLE   Antigens recognized by autologous antibody in patients with
             renal-cell carcinoma
  JOURNAL   Int. J. Cancer 83 (4), 456-464 (1999)
   MEDLINE   99438124
   PUBMED   10508479
REFERENCE   2 (bases 1 to 2817)
  AUTHORS   Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
             Jongeneel,V., Gure,A.O., Jager,D., Jager,E., Knuth,A., Chen,Y.-T.,
             and Old,L.J.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Ludwig Institute, Sloan-Kettering
             Institute, 1275 York Ave, New York, NY 10021, USA
FEATURES   Location/Qualifiers
     source   1..2817
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="1973/10-4"
                     /cell_type="renal cell carcinoma"
     CDS   50..1432
                     /codon_start=1
                     /product="putative protein kinase NY-REN-64 antigen"
                     /protein_id="AAD42884.1"
                     /db_xref="GI:5360131"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQGEGWKKLAVAIKKPSG
                     DDRYNQFHIRREALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQKQOMPFCDKDRTLMTPVQNLEOSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTEELKQPFQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVVMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     POLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSGASQCRHKKNKSPDIKKV
                     HQLLQEMTAS"
BASE COUNT   912 a   547 c   586 g   772 t
ORIGIN

US-10-001-254-26 (1-59) x AF155118 (1-2817)

Alignment Scores:
Pred. No.:
Score: 2.66e-30   Length: 2817
Percent Similarity: 290.00   Matches: 54
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 9   Indels: 0
                                        Gaps: 0

US-10-001-254-26 (1-59) x AF155118 (1-2817)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50 ATGAACAAACCCATAACACCATCAACATATGTGCGGCTGCTCAATGTTGGACTAATTAGG 109

Qy   21 LysLeuSerAspPheIleAspProGlnGlyGluGlyTrpLysLysLeuAlaValAlaIleLys 40
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   110 AAGCTGTCAGATTTTATTGATCCTCAAGGAGAAGGAGAGTTAGCTGTAGCTATTAAA 169

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
            ||||||||||||||||||||||||||||||||||||
Db   170 AAACCATCTGGTGATGATAGATACAATCAGTTCACATAAGG 211

Right column

KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
  AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
             Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
             Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2820)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
             Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
             University of Tokyo, Deprtment of Virology; Shirokane-dai, 4-6-1,
             Minato-ku, Tokyo 108-8639, Japan (E-mail:cdnal@ims.u-tokyo.ac.jp,
             Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
             International Trade and Industry of Japan; cDNA full insert
             sequencing; Research Association for Biotechnology; cDNA library
             construction, 5'- & 3'-end one pass sequencing; Department of
             Virology and Human Genome Center, Institute of Medical Science,
             University of Tokyo (partly supported by Science and Technology
             Agency).
FEATURES   Location/Qualifiers
     source   1..2820
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="KAT10395"
                     /cell_line="KATO III"
                     /cell_type="signet-ring cell carcinoma"
                     /clone_lib="KAT"
                     /note="cloning vector pME18SFL3"
     CDS   30..1412
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="BAA91232.1"
                     /db_xref="GI:7020684"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQGEGWKKLAVAIKKPSG
                     DDRYNQFHIRREALLQTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQKQOMPFCDKDRTLMTPVQNLEOSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTTVAVKKLAAM
                     VDITTEELKQPFQEIKVMAKCQHENLVELLGFSSDGDDLCLVYVVMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     POLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                     QOLLQEMTAS"
BASE COUNT   940 a   534 c   579 g   767 t
ORIGIN

US-10-001-254-26 (1-59) x AK000528 (1-2820)

Alignment Scores:
Pred. No.:
Score: 2.67e-30   Length: 2820
Percent Similarity: 290.00   Matches: 54
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 9   Indels: 0
                                        Gaps: 0

US-10-001-254-26 (1-59) x AK000528 (1-2820)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30 ATGAACAAACCCATAACACCATCAACATATGTGCGGCTGCTCAATGTTGGACTAATTAGG 89

Qy   21 LysLeuSerAspPheIleAspProGlnGlyGluGlyTrpLysLysLeuAlaValAlaIleLys 40
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   90 AAGCTGTCAGATTTTATTGATCCTCAAGGAGAAGGAGAGTTAGCTGTAGCTATTAAA 149

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
            ||||||||||||||||||||||||||||||||||||
...

RESULT 11
AK000528
LOCUS   AK000528   2820 bp   mRNA   linear   PRI 22-FEB-2000
DEFINITION Homo sapiens cDNA FLJ20521 fis, clone KAT10395.
ACCESSION   AK000528
VERSION   AK000528.1 GI:7020683

```
Db

RESULT 12
LOCUS           AY186092               33033 bp    DNA     linear   PRI 02-DEC-2002
DEFINITION      Homo sapiens interleukin-1 receptor-associated kinase 4 (IRAK4)
                gene, complete cds.
ACCESSION       AY186092
VERSION         AY186092.1  GI:26000792
KEYWORDS        .
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 33033)
  AUTHORS       Rieder,M.J., Armel,T.Z., Carrington,D.P., Ozuna,M., Kuldanek,S.A.,
                Rajkumar,N., Toth,E.J., Yi,Q. and Nickerson,D.A.
  TITLE         Direct Submission
  JOURNAL       Submitted (22-NOV-2002) Genome Sciences, University of Washington,
                1705 NE Pacific, Seattle, WA 98195, USA
COMMENT         To cite this work please use: SeattleSNPs. NHLBI HL66682 Program
                for Genomic Applications, UW-FHCRC, Seattle, WA (URL:
                http://pga.gs.washington.edu).
FEATURES             Location/Qualifiers
     source          1..33033
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     repeat_region   34..88
                     /rpt_family="L2"
     repeat_region   89..383
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       92..94
                     /replace=""
                     /rpt_type=dispersed
     variation       133
                     /gene="IRAK4"
                     /frequency="0.02"
                     /replace="c"
     repeat_region   384..599
                     /replace="a"
                     /frequency="0.06"
                     /rpt_family="L2"
                     /rpt_type=dispersed
     variation       642
                     /gene="IRAK4"
                     /frequency="0.02"
                     /replace="c"
     repeat_region   708..941
                     /rpt_family="L2"
                     /rpt_type=dispersed
     variation       854
                     /gene="IRAK4"
                     /frequency="0.02"
                     /replace="a"
     gene            1894..31033
                     /gene="IRAK4"
     mRNA            join(1894..1933,11040..11209,14157..14302,15117..15299,
                     15849..16009,16902..16966,20567..20681,21130..21239,
                     25244..25427,26599..26661,29336..29494,29617..31033)
                     /gene="IRAK4"
                     /product="interleukin-1 receptor-associated kinase 4"
     variation       1265
                     /gene="IRAK4"
                     /frequency="0.01"
                     /replace="c"
     variation       1533..1534
                     /gene="IRAK4"
                     /frequency="0.93"
                     /replace=""
     variation       2315
                     /gene="IRAK4"
                     /frequency="0.01"
                     /replace="c"
     variation       2355
                     /gene="IRAK4"
                     /frequency="0.01"
                     /replace="c"

     variation       2456
                     /gene="IRAK4"
                     /frequency="0.01"
                     /replace="t"
     repeat_region   2506..2807
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     variation       2940
                     /gene="IRAK4"
                     /frequency="0.04"
                     /replace="t"
     variation       3023
                     /gene="IRAK4"
                     /frequency="0.09"
                     /replace="t"
     repeat_region   3025..3915
                     /rpt_family="L2"
                     /rpt_type=dispersed
     variation       3038
                     /gene="IRAK4"
                     /frequency="0.01"
                     /replace="c"
     variation       3115..3129
                     /gene="IRAK4"
                     /frequency="0.72"
                     /replace=""
     variation       3189..
                     /gene="IRAK4"
                     /frequency="0.01"
                     /replace="t"
     variation       3536
                     /gene="IRAK4"
                     /frequency="0.14"
                     /replace="c"
     variation       3573
                     /gene="IRAK4"
                     /frequency="0.01"
                     /replace="g"
     variation       3819
                     /gene="IRAK4"
                     /frequency="0.01"
                     /replace="t"
     variation       3877
                     /gene="IRAK4"
                     /frequency="0.17"
                     /replace="t"
     variation       4144
                     /gene="IRAK4"
                     /frequency="0.04"
                     /replace="a"
     repeat_region   4438..4610
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     variation       4602
                     /gene="IRAK4"
                     /frequency="0.01"
                     /replace=""
     variation       4797
                     /gene="IRAK4"
                     /frequency="0.02"
                     /replace="a"
     variation       4972
                     /gene="IRAK4"
                     /frequency="0.31"
                     /replace=""
     repeat_region   5759..5797
                     /rpt_family="L2"
     variation       5821
                     /gene="IRAK4"
                     /frequency="0.02"
                     /replace="g"
     variation       5880
```

```
     variation      /gene="IRAK4"
                    /frequency="0.01"
                    /replace="c"
     5954
     repeat_region  /rpt_family="MER1_type"
     5964..6036     /rpt_type=dispersed
     variation      /gene="IRAK4"
                    /frequency="0.01"
                    /replace="t"
     repeat_region  6204 /rpt_family="MER1_type"
                    /rpt_type=dispersed
     variation      6204 /gene="IRAK4"
                    /frequency="0.01"
                    /replace=""
     repeat_region  6023..6590 /rpt_family="MER1_type"
                    /rpt_type=dispersed
     variation      /gene="IRAK4"
     6661..6662     /frequency="0.04"
                    /replace="g"
     variation      /gene="IRAK4"
     6698           /frequency="0.04"
                    /replace="t"
     variation      /gene="IRAK4"
     6998           /frequency="0.04"
                    /replace=""
     repeat_region  7074..7366 /rpt_family="Alu"
                    /rpt_type=dispersed
     repeat_region  7015..7073 /rpt_family="Mariner"
                    /rpt_type=dispersed
     variation      7122 /gene="IRAK4"
                    /frequency="0.04"
                    /replace="a"
     variation      7147 /gene="IRAK4"
                    /frequency="0.01"
                    /replace="a"
     variation      7234 /gene="IRAK4"
                    /frequency="0.01"
                    /replace="c"
     repeat_region  7367..8636 /rpt_family="Mariner"
                    /rpt_type=dispersed
     variation      7392 /gene="IRAK4"
                    /frequency="0.01"
                    /replace="c"
     variation      7597 /gene="IRAK4"
                    /frequency="0.01"
                    /replace="c"
     variation      7794 /gene="IRAK4"
                    /frequency="0.09"
                    /replace="g"
```

```
Alignment Scores:
Pred. No.:              3.32e-29          Length:        33033
Score:                  290.00            Matches:       54
Percent Similarity:     100.00%           Conservative:  0
Best Local Similarity:  89.78%            Mismatches:    0
Query Match:            9                 Indels:        0
                                          Gaps:          0

US-10-001-254-26 (1-59) x AY186092 (1-33033)

Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
Db 11049 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 11108
```

```
Qy    21 LysLeuSerAspPheIleAspProGlnGlyGlyTrpLysLeuLeuAlaValAlaIleLeuLys 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11109 AAGCTGTCAGATTTTATTGATCCTCAAGGAGGATGGAAGAGGTTAGCTGTAGCTATTAAA 11168

Qy    41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
         ||||||||||||||||||||||||||||||||||||||||||
Db 11169 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 11210
```

```
RESULT 13
AC093012

LOCUS       AC093012      118572 bp    DNA     linear   PRI 24-SEP-2002
DEFINITION  Homo sapiens 12 BAC RP11-210N13 (Roswell Park Cancer Institute
            Human BAC library) complete sequence.
ACCESSION   AC093012
VERSION     AC093012.9  GI:21629105
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 118572)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,N.C., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,J., Chiu,D., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J. Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
            Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
            Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
            Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
            Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Ioshikhes,I., Jackson,L.R., Jacobson,D., Jia,Y.,
            Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
            King,L., Korvah,J., Kovar,C., Kratovic,J. Kureshi,A., Landry,N.,
            Leal,B., Lee,E., Lewis,L.C., Lewis,T., Li,J., Li,Z., Lichtarge,O.,
            Liu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X.,
            Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
            Marondel,I., Martin,J., Martindale,A., Martinez,E., Massey,E.,
            Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
            Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
            Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
            Neal,D., Nelson,D., Newton,T., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Svetz,L., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.
REFERENCE   2  (bases 1 to 118572)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 118572)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
TITLE     Direct Submission
JOURNAL   Submitted (04-JUN-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4  (bases 1 to 118572)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (21-JUN-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5  (bases 1 to 118572)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 6  (bases 1 to 118572)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (03-SEP-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 7  (bases 1 to 118572)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (24-SEP-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jun 29, 2002 this sequence version replaced gi:21535906.
          INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
          gc-help@bcm.tmc.edu

          CLONE LENGTH: This sequence does not necessarily represent the
          entire insert of this clone. Overlapping regions of clones are only
          sequenced and submitted once, so the sequence for the remainder of
          the insert may be found in the record for the adjacent clones.
          Overlapping clones are noted at the beginning and end of the
          Features listing.

          ANNOTATION OF FEATURES:
          STSs are identified using ePCR (Genome Res. 7:541-550) searches
          of a local database that includes entries from dbSTS, GDB, and
          local mapping efforts.
          Repeats are identified using RepeatMasker (A. Smit and P. Green,
          unpublished.) for Human and Mouse sequences.
          Genes and Region of sequence similarity are identified by BLAST
          (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
          EST and cDNA sequences. Genes demonstrate at least two exons
          flanked by consensus splice sites that maintained sequence
          continuity across the splice junctions. Sequences that are not
          identical matches are annotated as similar.

          SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
          standard of double strand coverage with a minimum of 2 clones and 2
          reads with no ambiguities or 2 chemistries with a minimum of 2
          clones and 3 reads with no ambiguities. If the sequence quality for
          a region does not meet this standard, it will be indicated in the
          annotation as Low Coverage.

          QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
          standards - estimated error rate less than 1 per 10,000 bases.
          Reports of lowest quality individual bases and measures of base
          quality are listed below. Description of the metrics can be found
          at URL:
          http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
          ml.
FEATURES           Location/Qualifiers
     source        1..118572
                   /organism="Homo sapiens"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:9606"
                   /chromosome="12"
                   /clone="RP11-210N13"
```

```
     misc_feature     complement(1..2004)
                      /note="overlaps bases 1..2004 of clone AC016143"
                      /function="clone overlap"
     repeat_region    869..903
                      /rpt_family="HSMAR1"
     repeat_region    1047..1467
                      /rpt_family="MLT1K"
     repeat_region    1644..1760
                      /rpt_family="MIR"
     repeat_region    1985..2295
                      /rpt_family="AluY"
     repeat_region    2296..2325
                      /rpt_family="AT_rich"
     repeat_region    complement(2730..3024)
                      /rpt_family="AluSg"
     repeat_region    4092..4211
                      /rpt_family="MER5A"
     repeat_region    4389..4415
                      /rpt_family="AT_rich"
     repeat_region    4502..4561
                      /rpt_family="AT_rich"
     repeat_region    complement(4612..4778)
                      /rpt_family="MIR"
     repeat_region    5859..5926
                      /rpt_family="MLT1F"
     repeat_region    7574..8066
                      /rpt_family="AT_rich"
     repeat_region    complement(8266..8438)
                      /rpt_family="MIR"
     repeat_region    8889..8927
                      /rpt_family="MER5B"
     repeat_region    complement(11739..11889)
                      /rpt_family="AT_rich"
     repeat_region    11910..12124
                      /rpt_family="MER58A"
     repeat_region    12249..12617
                      /rpt_family="THE1C"
     repeat_region    14290..14310
                      /rpt_family="AT_rich"
     repeat_region    14463..14670
                      /rpt_family="AluSq"
     repeat_region    complement(14742..15039)
                      /rpt_family="AluSx"
     repeat_region    15669..15691
                      /rpt_family="(TTG)n"
     repeat_region    complement(15692..15972)
                      /rpt_family="AluSc"
     repeat_region    16265..16352
                      /rpt_family="L1PA16"
     repeat_region    16353..16631
                      /rpt_family="AluJo"
     repeat_region    16632..17007
                      /rpt_family="L1PA16"
     repeat_region    17329..17356
```

```
Alignment Scores:
Pred. No.:      1.23e-28           Length:       118572
Score:          290.00             Matches:      54
Percent Similarity:  100.00%       Conservative: 0
Best Local Similarity: 100.00%     Mismatches:   0
Query Match:    89.78%             Indels:       0
DB:             9                  Gaps:         0
```

US-10-001-254-26 (1-59) x AC093012 (1-118572)

```
Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 35063 ATGAACAAACCCATAACACCATCAACACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 35122

Qy    21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleAlaLys 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 35123 AAGCTGTCAGATTTTATTGATCCTCAGGAAGGATGGAAGAAGTTAGCTGTAGCTATTGCAAAA 35182

Qy    41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
           ||||||||||||||||||||||||||||||||||||||||||
Db 35183 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAGG 35224
```

RESULT 14
LOCUS   AC021719/c
DEFINITION Homo sapiens chromosome 19 clone RP11-210N13 map 19, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC021719
VERSION   AC021719.3 GI:7408021
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

LOCUS   AC021719   165868 bp   DNA   linear   HTG 24-AUG-2002

REFERENCE   1 (bases 1 to 165868)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Unpublished
JOURNAL

REFERENCE   2 (bases 1 to 165868)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Bouknagalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE   3 (bases 1 to 165868)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Bouknagalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,T., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Peterson,N., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigillo,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT   On Apr 4, 2000 this sequence version replaced gi:6957781.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------------- Genome Center -----------------
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information --------------
            Center project name: L5945
            Center clone name: 210_N13
            -------------- Summary Statistics --------------
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 156294 bases at least Q40
            Consensus quality: 160271 bases at least Q30
            Consensus quality: 162203 bases at least Q20
            Insert size: 175000; agarose-fp
            Insert size: 164568; sum-of-contigs
            Quality coverage: 4.9 in Q20 bases; agarose-fp
            Quality coverage: 5.2 in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   1982: contig of 1982 bp in length
            *   1983   2082: gap of 100 bp
            *   2083   4169: contig of 2087 bp in length
            *   4170   4269: gap of 100 bp
            *   4270   8449: contig of 4180 bp in length
            *   8450   8549: gap of 100 bp
            *   8550   15400: contig of 6851 bp in length
            * 15401   15500: gap of 100 bp
            * 15501   20885: contig of 5385 bp in length
            * 20886   20985: gap of 100 bp
            * 20986   27150: contig of 6165 bp in length
            * 27151   27250: gap of 100 bp
            * 27251   36127: contig of 8877 bp in length
            * 36128   36227: gap of 100 bp
            * 36228   47299: contig of 11072 bp in length
            * 47300   47399: gap of 100 bp
            * 47400   63394: contig of 15995 bp in length
            * 63395   63494: gap of 100 bp
            * 63495   78310: contig of 14816 bp in length
            * 78311   78410: gap of 100 bp
            * 78411   93139: contig of 14729 bp in length
            * 93140   93239: gap of 100 bp
            * 93240   107624: contig of 14385 bp in length
            * 107625   107724: gap of 100 bp
            * 107725   123977: contig of 16253 bp in length
            * 123978   124077: gap of 100 bp
            * 124078   165868: contig of 41791 bp in length
FEATURES   Location/Qualifiers
     source   1..165868
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19"
                     /clone="RP11-210N13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1..1982
                     /note="assembly_fragment"

FEATURES   Location/Qualifiers
JOURNAL   Patent: WO 0177168-A 149 18-OCT-2001;
              CORIXA CORPORATION (US)
TITLE   Compositions and methods for the therapy and diagnosis of lung
              cancer
AUTHORS   Lodes,M.J., Wang,T., Mohamath,R. and Indrias,C.Y.
REFERENCE   1
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
SOURCE   Homo sapiens
              Homo sapiens (human)
KEYWORDS   .
VERSION   AX321132.1 GI:17904576
ACCESSION   AX321132
DEFINITION   Sequence 149 from Patent WO0177168.
LOCUS   AX321132   501 bp   DNA   linear   PAT 15-DEC-2001
RESULT 15
AX321132

US-10-001-254-26 (1-59) x AC021719 (1-165868)

Alignment Scores:
Pred. No.:   6.29e-28
Score: 286.00
Percent Similarity: 98.15%   Length: 165868
Best Local Similarity: 98.15%   Matches: 53
Query Match: 88.54%   Conservative: 0
DB: Mismatches: 1
                                              Indels: 0
                                              Gaps: 0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleIleArg 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114811 ATGAACAAAACCCATAACACATCAACATATGTGCGCTGCTTTAATGTTGGACTAATTAGG 114752

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaAlaIleLeuIleLys 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114751 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGAGATGGAAGAGTTAGCTGTAGCTATTAAA 114692

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
         ||||||||||||||||||||||||||||||||||||||||||
Db 114691 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 114650

US-10-001-254-26 (1-59) x AC021719 (1-165868)

BASE COUNT   50117 a   31341 c   32315 g   50785 t   1310 others
ORIGIN

misc_feature   2083..4169
                /note="assembly_fragment"
                clone_end:T7
                vector_side:right"
misc_feature   4270..8449
                /note="assembly_fragment"
misc_feature   8550..15400
                /note="assembly_fragment"
misc_feature   15501..20885
                /note="assembly_fragment"
misc_feature   20986..27150
                /note="assembly_fragment"
misc_feature   27251..36127
                /note="assembly_fragment"
misc_feature   36228..47299
                /note="assembly_fragment"
misc_feature   47400..63394
                /note="assembly_fragment"
misc_feature   63495..78310
                /note="assembly_fragment"
misc_feature   78411..93139
                /note="assembly_fragment"
misc_feature   93240..107624
                clone_end:SP6
                vector_side:right"
                /note="assembly_fragment"
misc_feature   107725..123977
                /note="assembly_fragment"
misc_feature   124078..165868
                /note="assembly_fragment"

---

source   1..501
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"

BASE COUNT   166 a   102 c   97 g   134 t   2 others
ORIGIN

US-10-001-254-26 (1-59) x AX321132 (1-501)

Alignment Scores:
Pred. No.:   2.97e-29
Score: 277.00
Percent Similarity: 98.18%   Length: 501
Best Local Similarity: 98.18%   Matches: 54
Query Match: 85.76%   Conservative: 0
DB: Mismatches: 0
                                              Indels: 1
                                              Gaps: 0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleIleArg 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ATGAACAAAACCCATAACACATCAACATATGTGCGCTGCTTTAATGTTGGACTAATTAGG 80

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaAlaIleLeuIleLys 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 140

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
         ||||||||||||||||||||||||||||||||||||||||||
Db   141 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 183

Search completed: January 18, 2004, 04:27:07
Job time : 1025.41 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 18, 2004, 01:11:43 ; Search time 576.135 seconds
                                                    (without alignments)
                                                    2488.940 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Title: US-10-001-254-26
Perfect score: 323
Sequence: 1 MNKPITPSTYVRCLNVGLIR...........KKPSGDDRYNQFHIRCCSQN 59

Scoring table:
  BLOSUM62
  Xgapop 10.0 , Xgapext 0.5
  Ygapop 10.0 , Ygapext 0.5
  Fgapop   6.0 , Fgapext 7.0
  Delop   6.0 , Delext   7.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152424_19723/app_query.fasta_1.1109
 -DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
 -DOCALIGN=200 -THR_SCORE=pct -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US10001254 @CGN_1_1.4382 @runat_16012004_152424_19723 -NCPU=6 -ICPU=3
 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
  1: em_estba:*
  2: em_esthum:*
  3: em_estin:*
  4: em_estmu:*
  5: em_gss_inv:*
  6: em_gss_pln:*
  7: em_estpl:*
  8: em_estro:*
  9: em_htc:*
 10: em_gss_mam:*
 11: em_gss_fun:*
 12: em_gss_hum:*
 13: gb_est4:*
 14: gb_est3:*
 15: gb_htc:*
 16: gb_est2:*
 17: gb_est1:*
 18: em_estfun:*
 19: em_estom:*
 20: em_gss_hum:*
 21: em_gss_inv:*
 22: em_gss_pln:*
 23: em_gss_vrt:*
 24: em_gss_fun:*
 25: em_gss_mam:*
 26: em_gss_pro:*
 27: em_gss_rod:*
 28: em_gss_phg:*
 29: gb_gss2:*
     em_gss_vrl:*
     gb_gss1:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	323	100.0	284	9	AA114228	AA114228 zn75g05.r
2	296	91.6	576	10	BP238344	BP238344 601904613
3	290	89.8	540	9	AL699213	AL699213 DKF2p686K
4	290	89.8	719	10	BF696981	BF696981 602130160
5	290	89.8	811	10	BG616438	BG616438 602642772
6	290	89.8	859	10	BG164491	BG164491 602342026
7	284	87.9	313	12	BM151935	BM151935 TCBAP1E11
8	282	87.3	265	9	AW436511	AW436511 76774 MAR
9	274	84.8	402	10	BE482619	BE482619 168463 BA
10	271	83.9	453	10	BB860349	BB860349 BB860349
11	271	83.9	503	14	CA538859	CA538859 C0272B01-
12	271	83.9	507	10	BB866698	BB866698 BB866698
13	271	83.9	524	9	AW106160	AW106160 um23h11.y
14	271	83.9	575	4	BX522921	BX522921 RZPD Mus
15	271	83.9	598	13	BQ552228	BQ552228 H4014C09-
16	271	83.9	610	10	BB660378	BB660378 BB660378
17	271	83.9	637	10	BB613447	BB613447 BB613447
18	271	83.9	638	14	BY721552	BY721552 BY721552
19	271	83.9	663	14	BY726858	BY726858 BY726858
20	271	83.9	676	10	BB613167	BB613167 BB613167
21	271	83.9	1161	11	AK020397	AK020397 Mus muscu
22	271	83.9	2481	11	AK028837	AK028837 Mus muscu
23	269	83.3	2810	11	AK029028	AK029028 Mus muscu
24	269	83.3	858	10	BF687921	BF687921 602066996
25	197	61.0	629	9	AL647125	AL647125 AL647125
26	192	59.4	520	10	BE132064	BE132064 db41h09.y
27	192	59.4	664	12	BU035962	BU035962 603381326
28	190	58.8	852	13	BU209111	BU209111 603950034
29	186	57.6	600	9	AJ453616	AJ453616 AJ453616
30	186	57.6	670	9	AJ447581	AJ447581 AJ447581
31	159	49.2	541	10	BG691069	BG691069 340084 BA
32	135	41.8	632	14	CA365604	CA365604 640775 NC
33	125	38.7	555	14	AW423082	AW423082 f168b07.y
34	125	38.7	743	14	CA474136	CA474136 AGENCOURT
35	111	34.4	318	14	CB940191	CB940191 IpCGJx14-
36	100	31.0	380	12	BI883790	BI883790 fn07c09.y
37	74	22.9	477	10	AW958112	AW958112 EST370182
38	73.5	22.8	584	10	BG078473	BG078473 H3028C02-
39	73.5	22.8	588	12	BM022353	BM022353 id59c09.y
40	73.5	22.8	884	12	BI854371	BI854371 603381326
41	71	22.0	508	12	BJ505938	BJ505938 BJ505938
42	71	22.0	605	12	BJ499048	BJ499048 BJ499048
43	69	21.4	631	12	BM138979	BM138979 Hc_ad_15D
44	68.5	21.2	778	14	CB236976	CB236976 AGENCOURT
45	68	21.1	803	28	BH443382	BH443382 BOGGL71TR

RESULT 1
AA114228

DNA
encoding seq 26

AA114228

LOCUS   AA114228   284 bp   mRNA   linear   EST 13-NOV-1996
DEFINITION   zn75g05.r1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
             cDNA clone IMAGE:564056 5', mRNA sequence.
ACCESSION   AA114228
VERSION   AA114228.1 GI:1668121
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 284)

ALIGNMENTS

AUTHORS   Hiller,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
           Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
           ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
           ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
           Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
           Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE   Generation and analysis of 280,000 human expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   97044478
PUBMED   8889549
COMMENT   Contact: Wilson RK
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           Seq primer: -28M13 rev2 from Amersham.
FEATURES   Location/Qualifiers
     source   1..284
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:4595579"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:564056"
                     /tissue_type="neuroepithelial cells"
                     /dev_stage="Ntera-2 neuroepithelial cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /clone_lib="Stratagene NT2 neuronal precursor 937230"
                     /note="Organ: brain; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                     oligo dT. Uninduced, exponentially growing neuroepithelial
                     cells (Ntera-2/c1.D1). Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector: ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3'. ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTT 3'."
BASE COUNT   93 a   63 c   56 g   71 t   1 others
ORIGIN

Alignment Scores:
Pred. No.:   2.76e-36
Score: 323.00   Length: 284
Percent Similarity: 100.00%   Matches: 59
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 9   Indels: 0
                                                    Gaps: 0

US-10-001-254-26 (1-59) x AA114228 (1-284)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg   20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAACAAACCCATAACACCATCAACATATGTGGGCTGCCTCAAGTGTGGGACTAATTAGG 120

Qy   21 LysLeuSerAspPheIleAspProGlnGlyTrpLysLeuAlaValAlaIleAlaIleLys   40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AAGCTGTCAGATTTTATTGATCCTCAAGGAAGGATGGAAGTTAGCTGTAGCTATTAAA 180

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysSerGlnAsn 59
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAACCATCTGGTGATGATAGATACAATCAGTTCACATAAGATGCTGTTCCAAAAC 237

AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
           National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished
JOURNAL   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
COMMENT   Tissue Procurement: ATCC
           cDNA Library Preparation: CLONTECH Laboratories, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1032 row: i column: 03
           High quality sequence start: 6
           High quality sequence stop: 536.
FEATURES   Location/Qualifiers
     source   1..576
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4132682"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /tissue_type="from chronic myelogenous leukemia"
                     /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT   191 a   108 c   132 g   145 t
ORIGIN

Alignment Scores:
Pred. No.:   5.43e-32
Score: 296.00   Length: 576
Percent Similarity: 100.00%   Matches: 54
Best Local Similarity: 98.18%   Conservative: 0
Query Match: 91.64%   Mismatches: 1
DB: 10   Indels: 0
                                                    Gaps: 0

US-10-001-254-26 (1-59) x BF238344 (1-576)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg   20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 ATGAACAAACCCATAACACCATCAACATATGTGGGCTGCCTCAAGTGTGGGACTAATTAGG 127

Qy   21 LysLeuSerAspPheIleAspProGlnGlyTrpLysLeuAlaValAlaIleAlaIleLys   40
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   128 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGTTAGCTGTAGCTATTAAA 187

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCys 55
              |||||||||||||||||||||||||||||||||||||||||||||
Db   188 AAACCATCTGGTGATGATAGATACAATCAGTTCACATAAGATGC 232

RESULT 2
LOCUS   BF238344   576 bp   mRNA   linear   EST 14-NOV-2000
DEFINITION 601904613F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4132682 5',
            mRNA sequence.
ACCESSION   BF238344
VERSION   BF238344.1 GI:11152264
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

RESULT 3
LOCUS   AL699213   540 bp   mRNA   linear   EST 21-MAR-2002
DEFINITION DKFZp686K18112_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686K18112 5', mRNA sequence.
ACCESSION   AL699213
VERSION   AL699213.1 GI:19619753
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 540)
AUTHORS   Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann,S.).
TITLE   EST (Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann ,S.)
JOURNAL   Unpublished
COMMENT   Contact: Poustka A.J.
             Department Lehrach
             Max-Planck-Institute for Molecular Genetics
             Ihnestrasse 73, 14195 Berlin, Germany
             Tel: +49-30-8413623
             Fax: +49-30-8413128
             Email: poustka@mping-berlin-dahlem.mpg.de
             This is the 5' sequence of the clone insert
             Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
             sequenced by DKFZ (German Cancer Research Center,
             Heidelberg/Germany) within the cDNA sequencing consortium of the
             German Genome Project.
             No si sequence available.
             This clone (DKFZp686K18112) is available at the RZPD in Berlin.
             Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
             Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES   Location/Qualifiers
     source   1..540
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686K18112"
                     /tissue_type="human skeletal muscle"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"
BASE COUNT   150 a   104 c   111 g   174 t   1 others
ORIGIN

US-10-001-254-26 (1-59) x AL699213 (1-540)

Alignment Scores:
Pred. No.:   1
Score: 3.59e-31   Length: 540
Percent Similarity: 100.00%   Matches: 54
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 9   Mismatches: 0
DB: Indels: 0
                                         Gaps: 0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATGAACAAACCCATAACAACCATCAACATATGTCGCGTGCCTCAAATGTTGGACTAATTAGG 120

Qy   21 LysLeuSerAspPheIleAspProGlnGlyTrpLysLysLeuAlaValAlaIleLeuLys 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AAGCTGTCAGATTTTATTGATCCTCAAGGAAGGATGGAAGAAGTTAGCTGTGAGCTATTAA 180

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
        ||||||||||||||||||||||||||||||||||||||||||
Db 181 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 222

RESULT 4
BF696981
LOCUS   BF696981   719 bp   mRNA   linear   EST 22-DEC-2000
DEFINITION   602130160F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4287014 5',
             mRNA sequence.
ACCESSION   BF696981
VERSION   BF696981.1 GI:11182389
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 719)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.

TITLE
JOURNAL   Unpublished
COMMENT   National Institutes of Health, Mammalian Gene Collection (MGC)
             Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1124 row: g column: 15
             High quality sequence stop: 632.
FEATURES   Location/Qualifiers
     source   1..719
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4287014"
                     /tissue_type="primitive neuroectoderm"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_56"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgccctcggccc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCAGTTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGCGCCGAGCGGCCGCGGGCCCAATG-dT(30) BN-3',
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT   216 a   145 c   163 g   195 t
ORIGIN

US-10-001-254-26 (1-59) x BF696981 (1-719)

Alignment Scores:
Pred. No.:   1
Score: 5.34e-31   Length: 719
Percent Similarity: 100.00%   Matches: 54
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 10   Mismatches: 0
DB: Indels: 0
                                         Gaps: 0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   68 ATGAACAAACCCATAACAACCATCAACATATGTCGCGTGCCTCAAATGTTGGACTAATTAGG 127

Qy   21 LysLeuSerAspPheIleAspProGlnGlyTrpLysLysLeuAlaValAlaIleLeuLys 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 128 AAGCTGTCAGATTTTATTGATCCTCAAGGAAGGATGGAAGAAGTTAGCTGTGAGCTATTAAA 187

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
        ||||||||||||||||||||||||||||||||||||||||||
Db 188 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 229

RESULT 5
BG616438
LOCUS   BG616438   811 bp   mRNA   linear   EST 18-APR-2001
DEFINITION   602642772F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4773760 5',
             mRNA sequence.
ACCESSION   BG616438
VERSION   BG616438.1 GI:13667809
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 811)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)

Column 1

```
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: CLONETECH Laboratories, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1645 row: p column: 17
          High quality sequence start: 3
          High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1..811
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4773760"
                     /tissue_type="embryonal carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_61"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA, 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      246 a    196 c    164 g    205 t
ORIGIN

US-10-001-254-26 (1-59) x BG616438 (1-811)

Alignment Scores:
Pred. No.:              1
Score:                  6.31e-31         Length:         811
Percent Similarity:     290.00           Matches:        54
Best Local Similarity:  100.00%          Conservative:   0
Query Match:            89.78%           Mismatches:     0
DB:                     10               Indels:         0
                                         Gaps:           0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyTyrLeuIleArg 20
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  53 ATGAACAAACCATAACACCATCAACACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 112

Qy  21 LysLeuSerAspPheIleAspProGlnGlyIleTrpLysGlyLysLeuAlaValAlaAlaLeuLys 40
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 113 AAGCTGTCAGATTTTATTGATCCTCAAGGAAGGATCGGAAGAAGTTAGCTGTAGCTATTAAA 172

Qy  41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
       |||||||||||||||||||||||||||||||||||||||||
Db 173 AAACCATCTGGTGATGATAGATACAATCAGTTCACATAAGG 214

RESULT 6
BG616491
LOCUS       BG616491     859 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602342026F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4452055 5',
            mRNA sequence.
ACCESSION   BG616491
VERSION     BG616491.1  GI:12671194
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 859)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
```

Column 2

```
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM10240 row: d column: 08
          High quality sequence stop: 634.
FEATURES             Location/Qualifiers
     source          1..859
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4452055"
                     /tissue_type="hypernephroma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_89"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      266 a    168 c    190 g    235 t
ORIGIN

US-10-001-254-26 (1-59) x BG616491 (1-859)

Alignment Scores:
Pred. No.:              1
Score:                  6.83e-31         Length:         859
Percent Similarity:     290.00           Matches:        54
Best Local Similarity:  100.00%          Conservative:   0
Query Match:            89.78%           Mismatches:     0
DB:                     10               Indels:         0
                                         Gaps:           0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyTyrLeuIleArg 20
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35 ATGAACAAACCATAACACCATCAACACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 94

Qy  21 LysLeuSerAspPheIleAspProGlnGlyIleTrpLysGlyLysLeuAlaValAlaAlaLeuLys 40
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  95 AAGCTGTCAGATTTTATTGATCCTCAAGGAAGGATCGGAAGAAGTTAGCTGTAGCTATTAAA 154

Qy  41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
       |||||||||||||||||||||||||||||||||||||||||
Db 155 AAACCATCTGGTGATGATAGATACAATCAGTTCACATAAGG 196

RESULT 7
BM151935
LOCUS       BM151935     313 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  TCBAP1B11457 Pediatric pre-B cell acute lymphoblastic leukemia
            Baylor-HGSC project=TCBA Homo sapiens cDNA clone TCBAP1145, mRNA
            sequence.
ACCESSION   BM151935
VERSION     BM151935.1  GI:17175815
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 313)
  AUTHORS   Wei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman,F.R.Jr.,
            Gunaratne,P.H., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.
  TITLE     Pediatric Leukemia cDNA Sequencing Project (2001)
  JOURNAL   Unpublished
COMMENT   Contact: Dr. Judith F. Margolin
          Texas Children's Cancer Center and Human Genome Sequencing Center
          at Baylor College of Medicine
```

```
FEATURES
     source          1..313
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="TCBAP1145"
                     /sex="male"
                     /tissue_type="leukopheresis"
                     /cell_type="pre-B cell"
                     /dev_stage="pediatric 2 years"
                     /lab_host="DH10B"
                     /clone_lib="Pediatric pre-B cell acute lymphoblastic
                     leukemia Baylor-HGSC Project="TCBA"
                     /note="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
                     First strand cDNA was primed with an anchored
                     Xhol-oligo(dT) primer [5'GGAGGACTCGAGCCGCGGCCGAGGAGGAG(T)VN
                     3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand
                     was primed with a BamHI-dC primer
                     [5'AGAGGCTGGATCCCGCGCGGCCGCCAATAATAATAAT(C) 3'].
                     Double-stranded cDNA was then digested with BamHI and XhoI
                     and directionally cloned into the BamHI and SalI sites of
                     lambda pSB vector. Library went through one round of
                     normalization. Library was constructed by Wei Yu at RIKEN
                     of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
                     Itoh M, Nagaoka S, Sasakin, Okazaki Y, Muramatsu M,
                     Schneider C, Hayashizaki Y, High efficiency selection of
                     full-length cDNA by improved biotinylated cap trapper.,
                     DNA Res 4: 1, 61-6, Feb 28, 1997)"
BASE COUNT       82 a      62 c      81 g      87 t      1 others
ORIGIN

US-10-001-254-26 (1-59) x BM151935 (1-313)

Alignment Scores:
Pred. No.:              1.22e-30      Length:        313
Score:                  284.00        Matches:       53
Percent Similarity:     98.15%        Conservative:  0
Best Local Similarity:  98.15%        Mismatches:    1
Query Match:            87.93%        Indels:        0
DB:                     12            Gaps:          0

Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg 20
Db    1 ATGAACAAACCCCATAATACCTTCAACATATGTGCCGTGCTCCAARGTTGGACTAATTAGG 112

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLysLys 40
Db   53 ATGAACAAACCCCATAATACCTTCAACATATGTGCCGTGCTCCAARGTTGGACTAATTAGG 112

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
        ||||||||||||||||||||||||||||||||||||||||
Db  113 AAGCGTGTCAGATTTTATTGATCCTCAAGAAGGAGGATGGAAAGAAGTTAGCTGTAGCTATTAAA 172

Qy  173 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 214
        |||||||||||||||||||||||||||||||||||||||||
Db  173 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 214
```

```
RESULT 8
LOCUS       AW436511        265 bp      mRNA      linear      EST 09-JUL-2000
DEFINITION  76774 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   AW436511
VERSION     AW436511.1  GI:6971817
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
  AUTHORS   Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
            Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R., Quackenbush
            ,J. and Keele,J.W.
```

```
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smithe@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 31    row: H    column: 17
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
     source          1..265
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /clone="DH10B"
                     /lab_host="MARC 2PIG"
                     /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT       90 a      53 c      58 g      64 t
ORIGIN

US-10-001-254-26 (1-59) x AW436511 (1-265)

Alignment Scores:
Pred. No.:              1.87e-30      Length:        265
Score:                  282.00        Matches:       53
Percent Similarity:     98.15%        Conservative:  0
Best Local Similarity:  98.15%        Mismatches:    1
Query Match:            87.31%        Indels:        0
DB:                     9             Gaps:          0

Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg 20
Db   32 ATGAACAAACCCCATAATACAGCTTCAACATATGTGCCGTGCTCCAARGTTGGACTAATTAGG 91

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLysLys 40
        ||||||||||||||||||||||||||||||||||||||||||||
Db   92 AAGTTGTCAGATTTTATTGATCCTCAAGAAGGAGGATGGAAAGAAGTTAGCAGTAGCTATTAAA 151

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
        |||||||||||||||||||||||||||||||||||||||||
Db  152 AAACCATCTGGTGATGACAGATACAACCAGTTTCACATAAGG 193
```

```
RESULT 9
LOCUS       BE482619        402 bp      mRNA      linear      EST 27-MAR-2003
DEFINITION  160463 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE482619
VERSION     BE482619.1  GI:9602152
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE
  AUTHORS   Sonstegard,T., Capuco,A.V., White,J., Van Tassell,C.P., Connor,E.E.
            , Cho,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D. and
            Quackenbush,J.
TITLE       Analysis of bovine mammary gland EST and functional annotation of
            the Bos taurus gene index
```

1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@txccc.org
Seq primer: M13 primer.

```
FEATURES
     source          1..313
                     Location/Qualifiers
```

```
REFERENCE
  TITLE       Porcine gene discovery by normalized cDNA-library sequencing and
              EST cluster assembly
  JOURNAL     Mamm. Genome 13 (8), 475-478 (2002)
  MEDLINE     22213789
  PUBMED      12226715
```

JOURNAL   Mamm. Genome 13 (7), 373-379 (2002)
MEDLINE   22135956
PUBMED   12140684
COMMENT   Contact: Sonstegard TS
             USDA, ARS, Beltsville Agricultural Research Center
             Bdg. 200 Rm 2A, Beltsville, MD 20705, USA
             Tel: 301 504 8416
             Fax: 301 504 8414
             Email: tadb@lpsi.barc.usda.gov
             Single pass sequencing. Bases called and alt trimmed with phred
             v0.980904.e. Vector identified by cross_match with the -minscore 18
             and -minmatch 12 options.
             PCR Primers
             FORWARD: AGGAAACAGCTATGACCAT
             BACKWARD: GTTTTCCCAGTCACGACG
             Plate: 11 row: F column: 7
             Seq primer: ATTTAGGTGACACTATAG.
FEATURES   Location/Qualifiers
     source   1..402
                 /organism="Bos taurus"
                 /mol_type="mRNA"
                 /db_xref="taxon:9913"
                 /tissue_type="pooled"
                 /lab_host="DH10B"
                 /clone_lib="BARC 5BOV"
                 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
                 library made from pooled mRNA isolated from mammary
                 tissues at eight physiological, developmental, and disease
                 states."
BASE COUNT   114 a   95 c   88 g   105 t .
ORIGIN

Alignment Scores:
Pred. No.:
Score: 4.67e-29   Length: 402
Percent Similarity: 98.15%   Matches: 51
Best Local Similarity: 94.44%   Conservative: 2
Query Match: 84.83%   Mismatches: 1
DB: 10   Indels: 0
                                               Gaps: 0

US-10-001-254-26 (1-59) x BE482619 (1-402)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db   64 ATGAACAAAACCATAACAGCATCAACATACGTGCCCGTGCCTCAGTTCTTGGACTAATTAGA 123

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysGlyLeuAlaValAlaIleLeuLys 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 AAGCTGTCCGATTTTATTGATCCTCAAGAAGGATGGAAGAGTTAGCAGTCGCTATTATAAA 183

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
           |||||||||||||||||||||||||||||||||||||||||||
Db   184 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 225

TITLE   RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
JOURNAL   ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
             , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
             RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
             2001)
COMMENT   Unpublished
             Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center(GSC), Yokohama Institute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.go.jp.
             URL:http://genome.gsc.riken.go.jp/
             Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
             ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new
             genes. Genome Res. 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
             Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T. Matsuura
             ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
             Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer. Genome Res.
             10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
             ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
             encyclopedia: real-time sequence clustering for construction of a
             nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
             further details.
FEATURES   Location/Qualifiers
     source   1..453
                 /organism="Mus musculus"
                 /mol_type="mRNA"
                 /db_xref="taxon:10090"
                 /clone="G43001 3H14"
                 /clone_lib="RIKEN full-length enriched, pooled cell lines"
                 /note="pooled cell lines ; (cell_line=RCB-1751 WEHI 164),
                 (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                 (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                 (cell_line=RCB-0559 K-1), (cell_line=RCB-1283 B16
                 melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
                 ), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
                 (cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
                 (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                 (tissue_type=bone marrow, cell_type=stroma cell,
                 cell_line= CRL-2028 SR-4987), (tissue_type=colon,
                 cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                 cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                 cell_line=CRL-1734 SCA-9 clone I5), (strain=BALB/C,
                 cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3
                 ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                 BCH1)"
BASE COUNT   110 a   121 c   133 g   89 t .
ORIGIN

Alignment Scores:
Pred. No.:
Score: 1.48e-28.   Length: 453
Percent Similarity: 98.15%   Matches: 49
Best Local Similarity: 90.74%   Conservative: 4
Query Match: 83.90%   Mismatches: 1
DB: 10   Indels: 0
                                               Gaps: 0

US-10-001-254-26 (1-59) x BB860349 (1-453)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||:||||:|:|||
Db   165 ATGAACAAGACCGTTGCACCATCGACATACACCAACCTTAATGTGGGGGATCCTTAGG 224

RESULT 10
LOCUS   BB860349
DEFINITION   BB860349 RIKEN full-length enriched, pooled cell lines Mus musculus
             cDNA clone G43001 3H14 5', mRNA sequence.
ACCESSION   BB860349
VERSION   BB860349.1 GI:17101803
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 453)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
             Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
             ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
             Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
             , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
             Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa

```
Qy      21  LysLeuSerGlyAspAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleIleLys  40
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     225  AAGCTGTCGGGAGATTTAATTGATCCTCAAGAAGGGTGGAAGAAAATTAGCAGTAGCTATCAA  284

Qy      41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg  54
                ||||||||||||||||||||||||||||||||||||||||||
Db     285  AAGCCGTCCGGCGACGACAGATACAATCAGTTCCATATAAGG  326
```

RESULT 11
LOCUS   CA538859
DEFINITION C0272B01-5N NIA Mouse 7.5-dpc Whole Embryo cDNA Library (Long) Mus
            musculus cDNA clone NIA:C0272B01 IMAGE:30017484 5', mRNA sequence.
ACCESSION   CA538859
VERSION   CA538859.1 GI:25080828
KEYWORDS   EST.
SOURCE   Mus musculus
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 503)
  AUTHORS   Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Tanaka,T., Lim,M.K.,
            Luo,A. and Ko,M.S.H.
  TITLE   Systematic Analyses of NIA Mouse 7.5-dpc Whole Embryo cDNA Library
            (Long)
  JOURNAL   Unpublished
COMMENT   Other ESTs: C0272B01-3
            Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: C0272 row: B column: 01
            Seq primer: M13 Reverse
            High quality sequence stop: 503
            POLYA=No.
FEATURES   Location/Qualifiers
     source   1..503
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:C0272B01-5N"
                     /clone="NIA:C0272B01 IMAGE:30017484"
                     /tissue_type="whole embryo including extraembryonic
                     tissues at 7.5-days postcoitum"
                     /dev_stage="7.5-days postcoitum"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse 7.5-dpc Whole Embryo cDNA Library
                     (Long)"
                     /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11541199]). Total RNAs were
                     extracted from a pool of four embryos at 7.5-days
                     postcoitum. Double-stranded cDNAs were synthesized with an
                     oligo(dT) primer (Invitrogen):
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCCTTTTTTTTTTTTTTTTT-3') from
                     7 ug of total RNA, treated with T4 DNA polymerase, and
                     purified by ethanol-precipitation. The cDNAs were ligated
                     to Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                     The DH10B E. coli host was transformed with the ligation
                     mixture by the standard chemical method. The average
                     insert size is about 2.2 kb. The library was constructed

BASE COUNT   119 a   146 c   138 g   100 t
ORIGIN   by Yulan Piao (NIA)."

Alignment Scores:
  Pred. No.:   1.71e-28   Length: 503
  Score: 271.00   Matches: 49
  Percent Similarity: 98.15%   Conservative: 4
  Best Local Similarity: 90.74%   Mismatches: 1
  Query Match: 83.90%   Indels: 0
  DB: 14   Gaps: 0
```

US-10-001-254-26 (1-59) x CA538859 (1-503)

```
Qy       1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg  20
                ||||||||||||||||||||||||:|||:|||||||||||:::||||||:||||||||
Db     168  ATGAACAAGCCCGTTGACACATACAGTACATACGCAACCTTAATGTGGGAGGATCCTTAGG  227

Qy      21  LysLeuSerGlyAspAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleIleLys  40
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     228  AAGCTGTCGGGAGATTTAATTGATCCTCAAGAAGGGTGGAAGAAAATTAGCAGTAGCTATCAA  287

Qy      41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg  54
                ||||||||||||||||||||||||||||||||||||||||||
Db     288  AAGCCGTCCGGCGACGACAGATACAATCAGTTCCATATAAGG  329
```

RESULT 12
LOCUS   BB866698
DEFINITION BB866698 RIKEN full-length enriched, pooled cell lines Mus musculus
            cDNA clone G4D0001K20 5', mRNA sequence.
ACCESSION   BB866698
VERSION   BB866698.1 GI:17112908
KEYWORDS   EST.
SOURCE   Mus musculus
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 507)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,Y., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,Y., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Taku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res.   10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
```

## RESULT 13 (left column)

```
FEATURES             Location/Qualifiers
     source          1..507
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone_lib="RIKEN full-length enriched, pooled cell lines"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
                     (cell_line=RCB-0559 K-1, F1), (cell_line=CRL-1702 WEHI 231
                     (cell_type=B cells, cell_line=CRL-1283 B16
                     melanoma), (cell_type=B cells, cell_line=CRL-2065 MLTC-1),
                     , (cell_type=Leydig cells, cell_line=CRL-2070 NE),
                     (cell_type=Nullipotent stem cell, cell_line=CRL-1669 BCL1
                     (tissue_type=bladder, cell_line=RCB-0544 MBT-2),
                     (tissue_type=bone marrow, cell_line=RCB-1669 BCL1 Clone 13.20-3B3
                     ), (strain=C3H, tissue_type=stroma cell,
                     cell_line=CRL-1734 SCA-9, clone 15), (strain=BALB/C,
                     cell_line=CRL-2028 SR-4987), (tissue_type=colon,
                     cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
                     cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
                     cell_type=B cells, cell_line=CRL-1443
                     ), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
                     BC3H1)"
BASE COUNT           117 a    144 c    145 g    101 t
ORIGIN

US-10-001-254-26 (1-59) x BB866698 (1-507)

Alignment Scores:
Pred. No.:
Score:                          1.73e-28
Percent Similarity:             98.15%            Length:              507
Best Local Similarity:          90.74%            Matches:             49
Query Match:                    83.90%            Conservative:        4
DB:                             10                Mismatches:          1
                                                  Indels:              0
                                                  Gaps:                0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
        |||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db 187 ATGAACAAGCCGTTGACACATACCAGCACATACGCAACCTTAATGTGGGAGATCCTTAGG 246

Qy  21 LysLeuSerAspPheIleAspProGlnGluGlyTyrProLysLeuSerLeuAlaValAlaIleAlaIleLeuLys 40
        ||||||||||||||||||||||||||||||||||||||||||||
Db 247 AAGCTGTCGGATTTTATTGATCCTCAAGAAGGGGTGGAAGAAATTAGCAGTAGCTATCAAA 306

Qy  41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
        |||||||||||||||||||||||||||||||||||||
Db 307 AAGCCGTCCGGCGACGACAGATACAAATCAGTTCCATATAAGG 348
```

```
LOCUS           AW106160           524 bp    mRNA    linear    EST 20-OCT-1999
DEFINITION      um23h11.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
                IMAGE:2225445 5', mRNA sequence.
ACCESSION       AW106160
VERSION         AW106160.1  GI:6076896
KEYWORDS        EST.
SOURCE          Mus musculus
  ORGANISM      Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1  (bases 1 to 524)
  AUTHORS       Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
                Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
                ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
                ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
                Waterston,R. and Wilson,R.
  TITLE         The WashU-NCI Mouse EST Project 1999
```

RESULT 14 (right column)

```
  JOURNAL       Unpublished
  COMMENT       Other ESTs: um23h11.x1
                Contact: Marra M/WashU-NCI Mouse EST Project 1999
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: mouseest@watson.wustl.edu
                This clone is available royalty-free through LLNL ; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                MGI:1005641
                Seq primer: custom primer used
                High quality sequence stop: 510.
FEATURES             Location/Qualifiers
     source          1..524
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2225445"
                     /dev_stage="embryo, 14 dpc"
                     /clone_lib="Sugano mouse embryo mewa"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCAGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer [ATGTGGCCCTTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCAGTG). XhoI should be used to isolate the cDNA
                     insert.    Size selection was performed to exclude fragments
                     <1.5kb.    Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science).
                     Custom primers for sequencing: 5' end primer
                     CTTCTGCGTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA."
BASE COUNT           122 a    158 c    140 g    103 t    1 others
ORIGIN

US-10-001-254-26 (1-59) x AW106160 (1-524)

Alignment Scores:
Pred. No.:
Score:                          1.81e-28
Percent Similarity:             98.15%            Length:              524
Best Local Similarity:          90.74%            Matches:             49
Query Match:                    83.90%            Conservative:        4
DB:                             9                 Mismatches:          1
                                                  Indels:              0
                                                  Gaps:                0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
        |||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db 191 ATGAACAAGCCGTTGACACATACCAGCACATACGCAACCTTAATGTGGGAGATCCTTAGG 250

Qy  21 LysLeuSerAspPheIleAspProGlnGluGlyTyrProLysLeuSerLeuAlaValAlaIleAlaIleLeuLys 40
        |||||||||||||||||||||||||||||||||||||||||||
Db 251 AAGCTGTCGGATTTTATTGATCCTCAAGAAGGGGTGGAAGAAATTAGCAGTAGCTATCAAA 310

Qy  41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
        |||||||||||||||||||||||||||||||||||||
Db 311 AAGCCGTCCGGCGACGACAGATACAAATCAGTTCCATATAAGG 352
```

```
RESULT 14
ID       BX522921        standard; RNA; EST; 575 BP.
XX
AC       BX522921;
XX
SV       BX522921.1
XX
DT       27-MAY-2003 (Rel. 75, Created)
DT       27-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE       RZPD Mus musculus cDNA clone IMAGp998L225504 = IMAGE:2225445 5' EST.
```

XX
KW   EST; expressed sequence tag.
XX
OS   Mus musculus (house mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-575
RA   Heil O., Ebert L., Neubert P., Peters M., Radelof U., Schneider D.,
RA   Korn B.;
RT   ;
RL   Submitted (28-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer
RL   Feld 580, D-69120 Heidelberg, Germany
XX
DR   RZPD; IMAGp998L225504.
CC   RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
CC   Mouse UnigeneSet - RZPD2 (RZPDLIB No.981)
CC   http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=981
CC   Contact: Ina Rolfs
CC   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
CC   Heubnerweg 6, D-14059 Berlin, Germany
CC   Tel: +49 30 3639 101
CC   Fax: +49 30 3269 111
CC   www.rzpd.de
CC   This clone is available royalty-free from RZPD;
CC   contact RZPD (clone@rzpd.de) for further information.
CC   Seq primer: sugF, Primer sequence: CTTCTGCTCTAAAAGCTGCG
XX
FH   Key   Location/Qualifiers
FH
FT   source   1..575
FT   /db_xref="taxon:10090"
FT   /note="1st strand cDNA was primed with an oligo(dT) primer
FT   [ATGTGGCCGTTTTTTTTTTTTTTTTTT]; double-stranded cDNA was
FT   ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and
FT   cloned into distinct DraIII sites of the pME18S-FL3 vector
FT   (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used
FT   to isolate the cDNA insert. Size selection was performed to
FT   exclude fragments <1.5kb. Library constructed by Dr. Sumio
FT   Sugano (University of Tokyo Institute of Medical Science).
FT   Custom primers for sequencing: 5' end primer
FT   CTTCTGCTCTAAAAGCTGCG and 3' end primer
FT   CGACCTGCAGCTCCAGCACA. REFERENCES: Suzuki, Y., Yoshitomo,
FT   K., Maruyama, K., Suyama, A., and Sugano, S. Construction
FT   and characterization of a full length-enriched and a 5' end
FT   enriched cDNA library. Gene 200, 149-156, 1997. Sasaki, Z.,
FT   Suzuki, Y., Watanabe, M., Imai, J., Shibui, A., Yoshida,
FT   K., Hata, H., Yamaguchi, R., Tateyama, S., and Sugano, S.
FT   Construction of mouse full length-enriched cDNA libraries
FT   by oligo-capping. DNA Research, submitted."
FT   /organism="Mus musculus"
FT   /clone="IMAGp998L225504"
FT   /clone_lib="Sugano mouse embryo newa"
FT   /dev_stage="embryo, 14 dpc"
FT   /lab_host="DH10B"
XX
SQ   Sequence 575 BP; 132 A; 172 C; 155 G; 114 T; 2 other;

Alignment Scores:
Pred. No.:   2.06e-28
Score: 271.00
Percent Similarity: 98.15%   Length: 575
Best Local Similarity: 90.74%   Matches: 49
Query Match: 83.90%   Conservative: 4
DB: 4   Mismatches: 1
                                     Indels: 1
                                     Gaps: 0

US-10-001-254-26 (1-59) x BX522921 (1-575)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
        |||||||||||||||||||||||||||||||||||||||||||||:::|||||||:|||
Db 216 ATGAACAAGCCGTTGACACCATCGACATACCGGAACCTTAATGTGGGGATCCTTAGG 275

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLys 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 276 AAGCTGTCGGATTTTATTGATCCTCAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAA 335

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
        ||||||||||||||||||||||||||||||||||||||||||
Db 336 AAGCCGTCCGGCGACGACAGATACAATCAGTTCCATATAAGG 377

BQ552228
BQ552228
LOCUS   BQ552228   598 bp   mRNA   linear   EST 20-JUN-2002
DEFINITION H4014C09-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
            H4014C09 5', mRNA sequence.
ACCESSION   BQ552228
VERSION   BQ552228.1 GI:21453114
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 598)
  AUTHORS   VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G., Martin
            ,P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T., Kargul,G.J.,
            Luo,A.G., Kelso,J., Hide,W. and Ko,M.S.H.
  TITLE   Assembly, verification, and initial annotation of NIA 7.4K mouse
            cDNA clone set
  JOURNAL   Genome Res. 12 (12), 1999-2003 (2002)
   PUBMED   12466305
  MEDLINE   22354164
COMMENT   Contact: Yong Qian
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/NIA_7_4k.html for details.
            Plate: H4014 row: C column: 09
            Seq primer: -21M13 Reverse
            High quality sequence stop: 598
            PolyA=No.
FEATURES   Location/Qualifiers
     source   1..598
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="niaEST:H4014C09-5"
                     /db_xref="taxon:10090"
                     /clone="H4014C09"
                     /sex="mixed"
                     /dev_stage="mixed"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This
                     clone is among a rearrayed set of 7,407 clones from more
                     than 20 cDNA libraries."
BASE COUNT   150 a   174 c   163 g   111 t
ORIGIN

Alignment Scores:
Pred. No.:   2.18e-28
Score: 271.00
Percent Similarity: 98.15%   Length: 598
Best Local Similarity: 90.74%   Matches: 49
Query Match: 83.90%   Conservative: 4
DB: 13   Mismatches: 1
                                     Indels: 0
                                     Gaps: 0

US-10-001-254-26 (1-59) x BQ552228 (1-598)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
        |||||||||||||||||||||||||||||||||||||||||||||:::|||||||:|||
Db 152 ATGAACAAGCCGTTGACACCATCGACATACCGGAACCTTAATGTGGGGATCCTTAGG 211

```
Qy     21  LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLeuAlaValAlaIleLys  40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212  AAGCTGTCGGATTTTATTGATCCTCAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAA  271

Qy     41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg  54
           ||||||||||||||||||||||||||||||||||||||||||
Db    272  AAGCCGTCCGGCGACGACAGATACAATCAGTTCCATATAAGG  313
```

Search completed: January 18, 2004, 06:06:52
Job time : 578.135 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 00:40:22 ; Search time 70.7617 Seconds
                                                   (without alignments)
                                                   2250.748 Million cell updates/sec

Title: US-10-001-254-26
Perfect score: 323
Sequence: 1 MNKPITPSTYVRCLNVGLIR.........KKPSGDDRYNQFHIRCCSQN 59

Scoring table: BLOSUM62
                  Xgapop 10.0 , Xgapext 0.5
                  Ygapop 10.0 , Ygapext 0.5
                  Fgapop   6.0 , Fgapext 7.0
                  Delop   6.0 , Delext   7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
  -MODEL=frame+_p2n.model -DEV=xlh
  -Q=/cgn2.1/USPTO_spool/US10001254/runat_16012004_152423_19704/app_query.fasta_1.1109
  -DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
  -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
  -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
  -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
  -USER=US10001254 @CCN 1.1_0 @runat_16012004 152423_19704 -NCPU=6 -ICPU=3
  -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
  -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
  -FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7 -YGAPOP=10

Database :   N_Geneseq_19Jun03:*
  1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
  2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
  3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
  4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
  5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
  6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
  7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
  8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
  9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
 10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
 11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
 12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
 13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
 14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
 15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
 16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
 17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
 18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
 19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
 20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
 21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
 22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
 23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
 24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
 25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	323	100.0	211	24	AAD40084	Human IRAK4 short
2	323	100.0	415	23	AAS76802	DNA encoding novel
3	323	100.0	2213	22	AAH13798	Human cDNA sequenc
4	290	89.8	833	21	AAA09319	Human cancer assoc
5	290	89.8	1383	22	AAD10197	Human interleukin-
6	290	89.8	1383	24	AAD40079	Human IRAK4 gene #
7	290	89.8	1668	23	AAS76805	DNA encoding novel
8	290	89.8	2817	24	AAD40085	Human IRAK4 gene #
9	277	85.8	501	21	AAD61608	Lung small cell ca
10	271	83.9	1542	22	AAD10198	Mouse interleukin-
11	247	76.5	294	24	AAD40074	Human IRAK4 DD (de
12	68	21.1	5170	21	AAA16683	Human secreted pro
13	68	21.1	349980	24	ABQ81849	Bifidobacterium lo
14	66	20.4	972	23	ABL27971	Drosophila melanog
15	66	20.4	3610	21	ABL27970	Drosophila melanog
16	65.5	20.3	1257	21	AAC33451	Arabidopsis thalia
17	65.5	20.3	4151	22	ABA06564	Human cDNA SEQ ID
18	65.5	20.3	4151	22	ABV83901	Human polynucleoti
19	64.5	20.0	2893	23	ABL25610	Drosophila melanog
20	64	19.8	410	22	AAK56325	Human immune/haema
21	64	19.8	1301	25	ABQ84283	Full length cDNA 1
22	64	19.8	2388	25	ABA05888	Human aminopeptida
23	64	19.8	2391	25	ABQ84286	Mouse DPP10 coding
24	64	19.8	2583	24	ABK83324	cDNA encoding huma
25	64	19.8	2767	25	ABQ75945	Human PMM encodin
26	64	19.8	2909	25	ABQ84287	Human DPP10 cDNA s
27	64	19.8	2909	25	ABQ84292	Mouse DPP10 full l
28	64	19.8	3238	24	ABA05887	Human aminopeptida
29	64	19.8	3550	25	ABQ84285	Human DPP10 cDNA s
30	63.5	19.7	206	22	AAS11008	Cone snail cDNA en
31	63	19.5	565	24	ABQ52190	Oligonucleotide fo
32	63	19.5	565	24	ABQ52191	Oligonucleotide fo
33	63	19.5	2944528	24	ABA03041	Listeria monocytog
34	62.5	19.3	200	22	AAS11011	Cone snail cDNA en
35	62	19.2	29417	23	ABL13008	Drosophila melanog
36	62	19.2	29417	23	ABA66305	Drosophila melanog
37	61.5	19.0	304	22	ABA33370	Human foetal liver
38	61.5	19.0	304	22	AAK14722	Probe #11836 for g
39	61.5	19.0	304	22	AAK40465	Human bone marrow
40	61.5	19.0	304	22	AAI21224	Human brain expres
41	61.5	19.0	304	22	AAI46496	Probe #11157 for g
42	61.5	19.0	304	22	AAI06933	Probe #15182 used
43	61.5	19.0	304	24	ABS40029	Probe #6924 used t
44	61.5	19.0	304	24	ABS14456	Human liver single
45	61.5	19.0	304	23	ABS28667	Human liver single

RESULT 1
AAD40084 .

ID   AAD40084 standard; DNA; 211 BP.
XX
AC   AAD40084;
XX
DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 short gene.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
OS   Homo sapiens.

ALIGNMENTS

```
XX
FH   Key              Location/Qualifiers
FT   CDS              1..180
FT                    /*tag=
FT                    /product= "Human IRAK4 short protein"
XX
PN   WO20024O680-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US44844.
XX
PR   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24864.
XX
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection and other diseases
PT   -
XX
PS   Claim 19; Page 195; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 short gene.
XX
SQ   Sequence 211 BP; 77 A; 41 C; 36 G; 57 T; 0 other;
```

```
Qy    21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysGlyLeuAlaValAlaIleLys 40
Db       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGGGACTGGCTGTGGCTATTAAA 120

Qy    41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysCysSerGlnAsn 59
Db       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AAACCATCTGGTGATGATAGATACAAATCAGTTTCACATAAGATGTTCCCAAAAC 177
```

```
RESULT 2
AAS76802
ID   AAS76802 standard; cDNA; 415 BP.
XX
AC   AAS76802;
XX
DT   13-FEB-2002 (first entry)
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12615.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 12606; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 415 BP; 144 A; 83 C; 78 G; 110 T; 0 other;
```

```
XX                                            .
DE   DNA encoding novel human diagnostic protein #12606.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
```

```
Alignment Scores:
Pred. No.:
Score:                     1.63e-40            Length:      211
Percent Similarity:        323.00              Matches:      59
Best Local Similarity:     100.00%             Conservative: 0
Query Match:               100.00%             Mismatches:   0
DB:                        24                  Indels:       0
                                               Gaps:         0
```

```
US-10-001-254-26 (1-59) x AAD40084 (1-211)

Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
Db    1 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCTCAATGTTGGACTAATTAGG 60
```

```
Alignment Scores:
Pred. No.:
Score:          4.04e-40            Length:    415
                323.00             Matches:    59
```

Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 23   Indels: 0
DB: Gaps: 0

US-10-001-254-26 (1-59) x AAS76802 (1-415)

```
Qy   1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  36 ATGAACAAACCCATAACAACCATCAACATATGYGCGTGCCTCAATGTTGGACTAATTAGG 95

Qy  21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLeuLys 40
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  96 AAGCTGTGCAGATTTAATGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTGGCTATTAAA 155

Qy  41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysCysSerGlnAsn 59
       |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 156 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGAGTGCTGTTCCCAAAAC 212
```

RESULT 3
ID   AAH13798 standard; cDNA; 2213 BP.
XX
AC   AAH13798;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:10742.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0018776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesising polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 10742; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2213 BP; 642 A; 488 C; 516 G; 567 T; 0 other;

Alignment Scores:
Pred. No.:   3.84e-39   Length: 2213
Score: 323.00   Matches: 59
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 23   Indels: 0
DB: Gaps: 0

US-10-001-254-26 (1-59) x AAH13798 (1-2213)

```
Qy   1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 751 ATGAACAAACCCATAACAACCATCAACATATGYGCGTGCCTCAATGTTGGACTAATTAGG 810

Qy  21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLeuLys 40
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 811 AAGCTGTGCAGATTTAATGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTGGCTATTAAA 870

Qy  41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysCysSerGlnAsn 59
       |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 871 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGAGTGCTGTTCCCAAAAC 927
```

RESULT 4
AAA09319
ID   AAA09319 standard; DNA; 833 BP.
XX
AC   AAA09319;
XX
DT   10-AUG-2000 (first entry)
XX
DE   Human cancer associated antigen precursor DNA, clone NY-REN-64.
XX
KW   renal cancer; cancer associated antigen precursor; diagnosis;
KW   cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200020587-A2.
XX
PD   13-APR-2000.
XX
PF   04-OCT-1999; 99WO-US22873.
XX
PR   05-OCT-1998; 98US-0166300.
PR   05-OCT-1998; 98US-0166350.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y, Gout I, Tureci O, Sahin U, Pfreundschuh M, Scanlan MJ;
PI   Stockert E, Chen Y, Old LJ, Jager E, Knuth A;
XX
DR   WPI; 2000-303774/26.
DR   P-PSDB; AAY92347.
XX
PT   Preventing, diagnosing and/or treating disorders associated with
PT   abnormal expression of human cancer associated antigens
XX
PS   Claim 57; Page 85; 121pp; English.
XX
FH   Key   Location/Qualifiers
FT   CDS   50..670
FT   /*tag= a

XX
CC AAA09310-20 are novel genes isolated by SEREX screening from a renal
CC cancer cell line 1973/10.4. The genes encode cancer associated antigen
CC precursors. These gene products are useful in methods for preventing,
CC diagnosing and/or treating disorders, especially cancer, associated with
CC abnormal expression of human cancer associated antigens. The method
CC comprises contacting a sample from a subject with an agent that
CC specifically binds to the nucleic acid molecule or expression product
CC (or fragment) complexed with a human leukocyte antigen (HLA) molecule
CC and determining the interaction between the agent and the nucleic acid
CC molecule or the expression product as a determination of the disorder.
XX
SQ Sequence 833 BP; 273 A; 155 C; 179 G; 226 T; 0 other;

US-10-001-254-26 (1-59) x AAA09319 (1-833)

Alignment Scores:
Pred. No.:   1.23e-34
Score: 290.00   Length: 833
Percent Similarity: 100.00%   Matches: 54
Percent Identity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 89.78%   Indels: 0
DB: 21   Gaps: 0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50 ATGAACAAAACCCATAACACCATCAACACATATGTGCGCGCTGCCTCAATGTTGGACTAATTAGG 109

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysGlyLeuAlaValAlaIleLeuLys 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   110 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 169

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
        |||||||||||||||||||||||||||||||||||||||||
Db   170 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 211

RESULT 5
ID   AAD10197 standard; cDNA; 1383 BP.
XX
AC   AAD10197;
XX
DT   24-SEP-2001 (first entry)
XX
DE   Human interleukin-1 (IL-1) receptor-associated kinase (IRAK)-4 cDNA.
XX
KW   Human; interleukin-1 receptor associated kinase-4; IRAK-4; cytostatic;
KW   IL; antibacterial; antiinflammatory; ophthalmological; vasotropic; OPD;
KW   chronic obstructive pulmonary disease; neuroprotective; chronic cough;
KW   adult respiratory disease syndrome; pulmonary fibrosis; asthma; ARDS;
KW   interstitial lung disease; allergic rhinitis; transplant rejection;
KW   autoimmune disease; rheumatoid arthritis; systemic lupus erythematosus;
KW   multiple sclerosis; diabetes; cancer; solid tumour; lymphoma; stroke;
KW   cardiovascular disease; atherosclerosis; neurodegenerative disease;
KW   sepsis; osteoarthritis; rash; osteoporosis; psoriasis; dermatitis;
KW   inflammatory bowel disease; gout; Crohn's disease; ulcerative colitis;
KW   Behcet's syndrome; ankylosing spondylitis; IL-1 receptor/Toll receptor;
KW   sarcoidosis; transgenic animal; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1383
FT   /tag= a
FT   /product= "Human IRAK-4"
XX
PN   WO200151641-A1.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US01171.
XX
PR   13-JAN-2000; 2000US-0176395.

XX
PA   (TULA-) TULARIK INC.
XX
PI   Wesche H, Li S;
XX
DR   WPI; 2001-451860/48.
DR   P-PSDB; AAB05398.
XX
PT   Novel human interleukin-1 receptor associated kinase polypeptide,
PT   useful for identifying modulators of the polypeptide for treating gout,
PT   asthma, allergic rhinitis, multiple sclerosis and skin cancer -
XX
PS   Claim 7; Fig 2; 8pp; English.
XX
CC   The present sequence is a cDNA encoding human interleukin (IL)-1 receptor
CC   associated kinase (IRAK)-4. IRAK associate with activated IL-1, IL-18
CC   and other receptors and act to transduce signals originating from the
CC   activated receptors, ultimately leading to a variety of downstream
CC   effects such as nuclear factor (NF)-kappaB activation. The IRAK-4
CC   inhibitors are useful for treating inflammatory diseases such as
CC   pulmonary diseases and diseases of the airway (e.g., adult respiratory
CC   disease syndrome (ARDS), chronic obstructive pulmonary disease (OPD),
CC   pulmonary fibrosis, interstitial lung disease, asthma, chronic cough
CC   or allergic rhinitis), transplant rejection, autoimmune diseases (e.g.,
CC   rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis
CC   or diabetes), cancer (e.g., solid tumours, skin cancer or lymphoma),
CC   cardiovascular diseases (e.g., stroke and atherosclerosis), diseases
CC   of the central nervous system (e.g., neurodegenerative disease), CD14
CC   mediated sepsis, non-CD14 mediated sepsis, osteoarthritis, osteoporosis,
CC   psoriasis, diseases of the skin (e.g., rash, contact dermatitis, atopic
CC   dermatitis), inflammatory bowel disease (e.g., Crohn's disease and
CC   ulcerative colitis), Behcet's syndrome, ankylosing spondylitis, gout,
CC   sarcoidosis and ophthalmic diseases and conditions. The inhibitors of
CC   IRAK-4 activity or expression are used to inhibit signal transduction
CC   resulting from the activation of an interleukin-1 receptor (IL-1R)/Toll
CC   receptor in a cell. They also inhibit the activation of a transcription
CC   factor that activates NFkappaB in the cell. IRAK-4 is used to create a
CC   nonhuman transgenic animal which is useful for testing the function of
CC   IRAK-4 in vivo, to generate models for the study of inflammatory
CC   disorders and conditions and for the development of potential treatments
CC   for IRAK-4-related inflammatory diseases and conditions. IRAK-4 sequences
CC   are also used in gene therapy and in antisense therapy.
XX
SQ   Sequence 1383 BP; 463 A; 244 C; 283 G; 393 T; 0 other;

US-10-001-254-26 (1-59) x AAD10197 (1-1383)

Alignment Scores:
Pred. No.:   2.43e-34
Score: 290.00   Length: 1383
Percent Similarity: 100.00%   Matches: 54
Percent Identity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 89.78%   Indels: 0
DB: 22   Gaps: 0

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAACAAAACCCATAACACCATCAACACATATGTGCGCGCTGCCTCAATGTTGGACTAATTAGG 60

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysGlyLeuAlaValAlaIleLeuLys 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 120

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
        |||||||||||||||||||||||||||||||||||||||||
Db   121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 162

RESULT 6
ID   AAD40079 standard; DNA; 1383 BP.
XX
AC   AAD40079;
XX

DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 gene #1.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   immune-based pathology; allergy; autoimmunity; allograft rejection; cell division;
KW   inflammation; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200240680-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US44844.
XX
PR   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24859.
XX
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection and other diseases
PT   -
XX
PS   Claim 19; Page 180-182; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C.trachomatis , C.
CC   muridarum, C.pneumoniae, and C.psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 gene.
XX
SQ   Sequence 1383 BP; 463 A; 243 C; 283 G; 394 T; 0 other;

Alignment Scores:
Pred. No. :
Score: 2.43e-34   Length: 1383
                                290.00   Matches: 54

FH   Key   Location/Qualifiers
FT   CDS   1..1383
FT   /*tag= a
FT   /product= "Human IRAK4"

Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 89.78%   Indels: 0
DB: 24   Gaps: 0

US-10-001-254-26 (1-59) x AAD40079 (1-1383)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAACAAAAACCATACACACCATCAACACATATGTCGCGTGCCTCAATGTTGGACTAATAGG 60

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLeuAlaValAlaLeuLys 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGTTAGCTGTAGCTATTAAA 120

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
          |||||||||||||||||||||||||||||||||||||||||
Db 121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 162

RESULT 7
AAS76805
ID   AAS76805 standard; cDNA; 1668 BP.
XX
AC   AAS76805;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12609.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12618.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 12609; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (II) and (III) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and

RESULT 8

```
CC   amino acid sequences. AAS61197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1668 BP; 571 A; 289 C; 332 G; 476 T; 0 other;
```

US-10-001-254-26 (1-59) x AAS76805 (1-1668)

```
Alignment Scores:
Pred. No.:
Score:                    3.12e-34        Length:          1668
                          290.00          Matches:           54
Percent Similarity:       100.00%         Conservative:       0
Best Local Similarity:    100.00%         Mismatches:         0
Query Match:              89.78%          Indels:             0
                          23              Gaps:               0
```

```
Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGAACAAACCCATAACACCATCAACACATATGTGCGCTGCCTCAANGTTGGACTAATTAGG 60

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysGlySLeuAlaValAlaIleLeuLys 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 120

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
          ||||||||||||||||||||||||||||||||||||||||||
Db  121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 162
```

```
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection and other diseases
DR   P-PSDB; AAE24865.
DR   WPI; 2002-500222/53.
PR   29-JUN-2001; 2001US-301889P.
PR   17-NOV-2000; 2000US-0715893.
PF   15-NOV-2001; 2001WO-US44844.
PD   23-MAY-2002.
PN   WO0200240680-A2.
XX
PA   (BURN-) BURNHAM INST.
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
```

```
ID   AAD40085 standard; DNA; 2817 BP.
XX
AC   AAD40085;
XX
DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 gene #2.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             50..1432
FT                   /*tag= a
FT                   /product= "Human IRAK4"
XX
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
```

RESULT 9

```
CC   inflammation; allergy; autoimmunity; allograft rejection; cell division;
```

```
PT   -
XX
PS   Claim 19; Page 194-196; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis , C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 gene.
XX
SQ   Sequence 2817 BP; 912 A; 547 C; 586 G; 772 T; 0 other;
```

US-10-001-254-26 (1-59) x AAD40085 (1-2817)

```
Alignment Scores:
Pred. No.:
Score:                    6.32e-34        Length:          2817
                          290.00          Matches:           54
Percent Similarity:       100.00%         Conservative:       0
Best Local Similarity:    100.00%         Mismatches:         0
Query Match:              89.78%          Indels:             0
                          24              Gaps:               0
```

```
Qy    1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   50 ATGAACAAACCCATAACACCATCAACACATATGTGCGCTGCCTCAANGTTGGACTAATTAGG 109

Qy   21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysGlySLeuAlaValAlaIleLeuLys 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  110 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAA 169

Qy   41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
          ||||||||||||||||||||||||||||||||||||||||||
Db  170 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 211
```

```
RESULT 9
ID   AAS61608 standard; cDNA; 501 BP.
XX
AC   AAS61608;
XX
DT   29-JAN-2002 (first entry)
XX
DE   Lung small cell carcinoma antigen, cDNA #149.
XX
OS   Homo sapiens.
XX
KW   tumour; lung cancer; ss.
KW   Human; cytostatic; antitumour; lung small cell cancer antigen;
XX
PN   WO200171168-A2.
XX
PD   18-OCT-2001.
```

XX
PF   11-APR-2001; 2001WO-US11859.
XX
PR   11-APR-2000; 2000US-196780P.
PR   21-JUN-2000; 2000US-213361P.
PR   01-SEP-2000; 2000US-229763P.
PR   05-SEP-2000; 2000US-230629P.
PR   14-SEP-2000; 2000US-232565P.
PR   19-DEC-2000; 2000US-257037P.
PR   08-JAN-2001; 2001US-260796P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Lodes MJ, Wang T, Mohamath R, Indirias CY;
XX
DR   WPI; 2002-010896/01.
XX
PT   Lung tumour polynucleotide and polypeptides useful in therapy and
PT   diagnosis of cancer especially lung cancer -
XX
PS   Claim 1; Page 174; 295pp; English.
XX
XX
CC   The invention relates to novel isolated lung small cell cancer antigen
CC   polynucleotides (I) and polypeptides (II) used in a method of detecting
CC   cancer in a patient. The method is optionally performed by
CC   utilising oligonucleotides (III), where the biological sample
CC   from the patient is contacted with (III), detecting the amount of
CC   polynucleotide hybridised to (III) in the sample and comparing the
CC   amount of polynucleotide to a predetermined cut-off value and thereby
CC   determining cancer in a patient. (I), (II) or antigen-presenting cells
CC   specific for a tumour protein. The method comprises contacting T cells
CC   expressing (II) is useful for stimulating and/or expanding T cells
CC   with one of the components under conditions to permit the stimulation
CC   and/or expansion of the cells. A composition comprising (I) is useful for
CC   stimulating an immune response in a patient and for inhibiting the
CC   development of a cancer especially lung cancer in a patient. An
CC   isolated T cell population is useful for removing tumour cells from the
CC   biological sample and for inhibiting the development of cancer in a
CC   patient. AA561460-AA561874 represent novel human lung small cell
CC   cancer antigen coding sequences of the invention.
XX
XX
SQ   Sequence 501 BP; 166 A; 102 C; 97 G; 134 T; 2 other;

Alignment Scores:
Pred. No.:
Score: 6.18e-33
                        277.00
Percent Similarity: 98.18%
Best Local Similarity: 98.18%
Query Match: 85.76%
DB: 24

Length: 501
Matches: 54
Conservative: 0
Mismatches: 0
Indels: 1
Gaps: 0

US-10-001-254-26 (1-59) x AAS61608 (1-501)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ATGAACAAACCCATAACACCATCAACATATGTGCGCTGCCTCAATGTTGGACTAATTAGG 80

Qy   21 LysLeuSerAspPheIleAspProGlnGlyLeuGlyTrpLysLeuSerLeuAlaValAlaIleAlaLys 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 AAGCTGTCAGATTTTATTGATCCTCAAGGACGATGGAAGAAGTTAGCTGTAGCTATTAAA 140

Qy   41 LysProSerGlyAspAspSerProAlaGlyTyrAsnGln-PheHisIleArg 54
          ||||||||||||||||||||||||||||||||||||||||
Db   141 AAACCATCTGGTGATGATAGTACAATCAAGTTTCACATAAGG 183

RESULT 10
AAD10198
ID   AAD10198 standard; cDNA; 1542 BP.
XX
AC   AAD10198;
XX
DT   24-SEP-2001 (first entry)
XX

DE   Mouse interleukin-1 (IL-1) receptor-associated kinase (IRAK)-4 cDNA.
XX
KW   Mouse; interleukin-1 receptor associated kinase-4; IRAK-4; cytostatic;
KW   IL; antibacterial; antiinflammatory; ophthalmological; vasotropic; OPD;
KW   chronic obstructive pulmonary disease; neuroprotective; chronic cough;
KW   adult respiratory disease syndrome; allergic rhinitis; asthma; ARDS;
KW   interstitial lung disease; allergic rhinitis; transplant rejection;
KW   autoimmune disease; rheumatoid arthritis; systemic lupus erythematosus;
KW   multiple sclerosis; diabetes; cancer; solid tumour; lymphoma; stroke;
KW   cardiovascular disease; atherosclerosis; neurodegenerative disease;
KW   sepsis; osteoarthritis; rash; osteoporosis; psoriasis; dermatitis;
KW   inflammatory bowel disease; gout; Crohn's disease; ulcerative colitis;
KW   Behcet's syndrome; ankylosing spondylitis; IL-1 receptor/Toll receptor;
KW   sarcoidosis; transgenic animal; ss.
XX
OS   Mus sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1542
FT   /*tag= a
FT   /product= "Mouse IRAK-4"
XX
PN   WO20015164-A1.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US01171.
XX
PR   13-JAN-2000; 2000WO-US0176395.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Wesche H, Li S;
XX
DR   WPI; 2001-451860/48.
DR   P-PSDB; AAE05399.
XX
PT   Novel human interleukin-1 receptor associated kinase polypeptide,
PT   useful for identifying modulators of the polypeptide for treating gout,
PT   asthma, allergic rhinitis, multiple sclerosis and skin cancer -
XX
PS   Claim 26; Fig 4; 89pp; English.
XX
XX
CC   The present sequence is a cDNA encoding mouse interleukin (IL)-1 receptor
CC   associated kinase (IRAK)-4. IRAK associate with activated IL-1, IL-18
CC   and other receptors and act to transduce signals originating from the
CC   activated receptors, ultimately leading to a variety of downstream
CC   effects such as nuclear factor (NF)-kappaB activation. The IRAK-4
CC   inhibitors are useful for treating inflammatory diseases such as
CC   pulmonary diseases and diseases of the airway (e.g., adult respiratory
CC   disease syndrome (ARDS), chronic obstructive pulmonary disease (OPD),
CC   pulmonary fibrosis, interstitial lung disease, asthma, chronic cough
CC   or allergic rhinitis), transplant rejection, autoimmune diseases (e.g.,
CC   rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis
CC   or diabetes), cancer (e.g., solid tumours, skin cancer or lymphoma),
CC   cardiovascular diseases (e.g., stroke and atherosclerosis), diseases
CC   of the central nervous system (e.g., neurodegenerative disease), CD14
CC   mediated sepsis, non-CD14 mediated sepsis, osteoarthritis, osteoporosis,
CC   psoriasis, diseases of the skin (e.g., rash, contact dermatitis, atopic
CC   dermatitis), inflammatory bowel disease (e.g., Crohn's disease and
CC   ulcerative colitis), Behcet's syndrome, ankylosing spondylitis, gout,
CC   sarcoidosis and ophthalmic diseases and conditions. The inhibitors of
CC   IRAK-4 activity or expression are used to inhibit signal transduction
CC   resulting from the activation of an interleukin-1 receptor (IL-1R)/Toll
CC   receptor in a cell. They also inhibit the activation of a transcription
CC   factor that activates NFkappaB in the cell. IRAK-4 is used to create a
CC   nonhuman transgenic animal which is useful for testing the function of
CC   IRAK-4 in vivo, to generate models for the study of inflammatory
CC   disorders and conditions and for the development of potential treatments
CC   for IRAK-4-related inflammatory diseases and conditions. IRAK-4 sequences
CC   are also used in gene therapy and in antisense therapy.
XX
XX
SQ   Sequence 1542 BP; 421 A; 392 C; 423 G; 306 T; 0 other;

9

```
Alignment Scores:
Pred. No.:                          Length:       1542
Score:                  271.00      Matches:      49
Percent Similarity:     98.15%      Conservative: 4
Best Local Similarity:  90.74%      Mismatches:   1
Query Match:            83.90%      Indels:       0
                                    Gaps:         0

US-10-001-254-26 (1-59) x AAD10198 (1-1542)

Qy    1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleLeuArg  20
            |||||||||||||:||||::|||||::||||||||||||||||||||||
Db  163  ATGAACAAGCCGTTCACACCATCGACAGTACATACCGAACCTTAATGTGGGGGATCCTTAAGG  222

Qy   21  LysLeuSerAspPheIleAspProGlnGlyGlyTyrTrpLysGlySerLeuAlaValAlaIleLeuLys  40
            ||||||||||||||||||||||||||||||||||||||||
Db  223  AAGCTGTCGGATTTTATTGATCCTCAAGGAGGCTGGAAGAAATTAGCAGTAGCTATACAAA  282

Qy   41  LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg  54
            ||||||||||||||||||||||||
Db  283  AAGCCGTCCCGCGACGACAGATACAATCAGTTCCATATAAGG  324
```

RESULT 11

```
AC   AAD40074;
XX
ID   AAD40074 standard; cDNA; 294 BP.
XX
DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 DD (death domain) cDNA.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             1..294
FT                   /*tag= a
FT                   /product= "Human IRAK4 DD"
FT                   /note= "No start and stop codon"
FT                   /partial
XX
```

```
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection, cell division, B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis, C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathology. The protein
CC   is useful for treating immune-based pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The present sequence is human
CC   therapy and gene therapy. The invention is human IRAK4 DD cDNA.
XX
SQ   Sequence 294 BP; 84 A; 55 C; 62 G; 93 T; 0 other;
```

```
Alignment Scores:
Pred. No.:                          Length:       294
Score:                  247.00      Matches:      46
Percent Similarity:     100.00%     Conservative: 0
Best Local Similarity:  100.00%     Mismatches:   0
Query Match:            76.47%      Indels:       0
                                    Gaps:         0

US-10-001-254-26 (1-59) x AAD40074 (1-294)

Qy    9  ThrTyrValArgCysLeuAsnValGlyLeuIleLeuArgGlyLeuSerAspAspPheIleAspPro  28
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ACATACGTGGCGCTGCCTCAATGTTGGACTAATTACGAAGCTGTCAGATTTTATTGATCCT  60

Qy   29  GlnGlyGlyTyrTrpLysGlySerLeuAlaValAlaIleLeuLysLysProSerGlyAspAspArgTyr  48
            ||||||||||||||||||||||||||||||||||||||||
Db   61  CAAGGAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATTCGGTGATGATAGATAC  120

Qy   49  AsnGlnPheHisIleArg  54
            ||||||||||||||||||
Db  121  AATCAGTTTCACATAGG  138
```

RESULT 12

```
AC   AAA16683;
XX
ID   AAA16683 standard; cDNA; 5170 BP.
XX
DT   16-JUN-2000 (first entry)
XX
DE   Human secreted protein clone nf56_3 nucleotide sequence SEQ ID NO:131.
XX
KW   Human; secreted protein; immunostimulant; immunosuppressant; virucide;
KW   antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW   antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW   antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW   infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW   connective tissue disease; multiple sclerosis; erythematosis;
KW   rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW   Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW   insulin dependent diabetes mellitus; graft-versus-host-disease;
KW   autoimmune inflammatory eye disease; allergy; ss.
XX
OS   Homo sapiens.
XX
```

New polypeptide comprising a death domain or death effector domain,
useful for discovery of drugs that suppress infection, inflammation,
allergy, sepsis, autoimmunity, allograft rejection and other diseases
-
Claim 18; Page 173-174; 209pp; English.

The invention relates to an isolated polypeptide comprising a death

PN   WO20009952-A1.
XX
PD   24-FEB-2000.
XX
PF   13-AUG-1999;
XX
PR   14-AUG-1998;   98US-0096622.
PR   17-AUG-1998;   98US-0096815.
PR   04-SEP-1998;   98US-0099229.
PR   23-OCT-1998;   98US-0105368.
PR   08-JAN-1999;   99US-0115234.
PR   12-FEB-1999;   99US-0119931.
PR   18-FEB-1999;   99US-0120575.
PR   30-APR-1999;   99US-0132020.
PR   11-AUG-1999;   99US-0096622.
XX
PA   (GEMX) GENETICS INST INC.
XX
PI   Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI   Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI   Wong GG, Clark HF, Fechtel K;
XX
PT   antiinflammatory or tumor inhibition activity -
PT   hematopoiesis regulating, tissue growth, activin/inhibin
PT   nutritional, chemokine, immune stimulating or suppressing,
PA   New polynucleotides encoding secreted proteins, which may have e.g.
XX
DR   WPI; 2000-205979/18.
DR   P-PSDB; AAY94963.
XX
PS   Claim 140; Page 590-592; 641pp; English.
XX
CC   AAA16618 to AAA16697 encode the human secreted proteins given in
CC   AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
CC   adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
CC   adult placenta, adult testis, whole embryo, adult cartilage, kidney,
CC   foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC   and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC   predicted to have biological activities which would make them suitable
CC   for treating, preventing or ameliorating medical conditions in humans
CC   and animals. The polynucleotides can be used as markers for tissues in
CC   which the protein is preferentially expressed, as molecular weight
CC   markers on Southern gels, and as chromosome markers or tags to identify
CC   chromosomes or to map gene positions. The proteins can be used in the
CC   treatment of immune deficiencies and disorders, such as severe combined
CC   immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC   infections. These infections include human immunodeficiency virus (HIV),
CC   hepatitis, herpesviruses, mycobacteria, leishmania spp., malaria and
CC   candidiasis. The proteins can be used to treat autoimmune disorders such
CC   as connective tissue disease, multiple sclerosis, systemic lupus
CC   erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC   Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC   diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC   autoimmune inflammatory eye disease. The proteins can also be used to
CC   treat allergic conditions, such asthma. AAA16698 to AAA16774 represent
CC   probes for the human secreted proteins from the present invention.
XX
SQ   Sequence 5170 BP; 1372 A; 1257 C; 1174 G; 1367 T; 0 other;

Alignment Scores:

Qy   US-10-001-254-26 (1-59) x AAA16683 (1-5170)

Pred. No. :
Score: 19.8
Percent Similarity: 68.00
Best Local Similarity: 46.34%
Percent Match: 34.15%
Query Match: 21.05%
DB:

Length: 5170
Matches: 14
Conservative: 5
Mismatches: 10
Indels: 12
Gaps: 1

Db   4523 ATAGGTCTCATTAAACAGAGTGGAGAGATTTCATTGGTCAGCACTCCTCAATGAAA------ 4576
Qy   16 ValGlyLeuIleArgLysSerAspPheIleAspProGlnGlyTrpLysLysLeu 35
        :::||||||.|||||||||.::|||||.||.||||
Db   4607 TGC 4609

Qy   36 AlaValAlaIleLysGlyLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCys 55
                                                            ||||:|||:|||||:||||
Db   4577 --------------------------------GACAGACCTAATGACTGGCATTTGAGATGC 4606

Qy   56 Cys 56
         |||
Db   4607 TGC 4609

RESULT 13
ABQ81849
ID   ABQ81849 standard; DNA; 349980 BP.
XX
AC   ABQ81849;
XX
DT   19-NOV-2002 (first entry)
XX
DE   Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1105.
XX
KW   Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW   antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW   identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW   rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
OS   Bifidobacterium longum.
OS   Synthetic.
XX
PN   EP1227152-A1.
XX
PD   31-JUL-2002.
XX
PF   30-JAN-2001; 2001EP-0102050.
XX
PR   30-JAN-2001; 2001EP-0102050.
XX
PA   (NEST) SOC PROD NESTLE SA.
XX
DR   WPI; 2002-668397/72.
XX
PS   Disclosure; SEQ ID 1105; 80pp; English.
XX
PT   longum in a biological sample -
PT   as a probe or primer for detecting and/or identifying Bifidobacterium
PT   Novel polynucleotide comprising Bifidobacterium genome sequence useful
XX
CC   The present invention describes a polynucleotide (I) comprising a
CC   sequence of a Bifidobacterium genome selected from the nucleotide
CC   sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC   least 90% identity or which hybridises with the sequences given in
CC   ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding
CC   a fusion protein, comprising a sequence selected from 1097 sequences
CC   given in ABP65258 to ABP66354 ligated in frame to a polynucleotide
CC   encoding a heterologous polypeptide. (II) has antidiarrheic and
CC   antibacterial activities, and can be used as an inhibitor of Salmonella.
CC   (I) (which is a probe) is useful for the detection and/or identification
CC   of Bifidobacterium longum in a biological sample. A carrier containing
CC   the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618)
CC   can be used for preventing and/or treating diarrhoea brought about by
CC   pathogenic bacteria and/or rotavirus. The carrier is a food composition
CC   selected from milk, yogurt, curd, cheese, fermented milks, milk based
CC   fermented products, ice-creams, fermented cereal based products, milk
CC   based powders, infant formula, pet food or a pharmaceutical composition
CC   selected from tables, liquid bacterial suspensions, dried oral
CC   supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC   (I) is useful in DNA arrays or chips to carry out analysis of the
CC   expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC   Bifidobacterium related nucleotide sequences given in the Sequence
CC   Listing from the present invention but not mentioned further within the
CC   specification.
CC   N.B. The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied by the
CC   European Patent Office.
XX
SQ   Sequence 349980 BP; 68046 A; 106490 C; 106389 G; 69055 T; 0 other;

Left column — RESULT 14

Alignment Scores:
Pred. No.:
Score: 5.73e+03   Length: 349980
Percent Similarity: 68.00   Matches: 19
Best Local Similarity: 61.36%   Conservative: 8
Query Match: 43.18%   Mismatches: 9
                         21.05%   Indels: 8
DB: 24   Gaps: 3

US-10-001-254-26 (1-59) x ABQ81849 (1-349980)

```
Qy     12 ArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspProGlnGlu--- 30
             |||||||||||| ||||||||||||  |||||||  |||  |   |
Db  71940 CGTTGCCTCGCATTGGGGCCTCAAGCGAGGAACCGGCCGATTATCTTGATCCACACAATAT 71999

Qy     31 ---GlyTrpIleLysLeuAlaValAlaIleAlaIleLeuLysProSer--------GlyAspAsp 46
             :::||||||:::|||   ::|||
Db  72000 ---GTACTGAAACCGTCATGATTAACGGTGAT 72050

Qy     47 ArgTyrAsnGln 50
             ||||
Db  72051 CGGCCGGGACAG 72062
```

RESULT 14
ID   ABL27971
     ABL27971 standard; DNA; 972 BP.
XX
AC   ABL27971;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 35386.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 35386; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL16176-ABL30511) and the encoded proteins
CC   (ABS57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 972 BP; 244 A; 283 C; 266 G; 179 T; 0 other;

Right column — RESULT 15

Alignment Scores:
Pred. No.:
Score: 4.23   Length: 972
Percent Similarity: 66.00   Matches: 19
Best Local Similarity: 52.73%   Conservative: 10
Query Match: 34.55%   Mismatches: 20
                         20.43%   Indels: 6
DB: 23   Gaps: 2

US-10-001-254-26 (1-59) x ABL27971 (1-972)

```
Qy      2 AsnLysProIleThrProSerThrTyrValArgCysLeuAsnVal--------------- 16
            |||||||| |||||| ||:::|::||  |||||||||:::
Db    280 AATCCTCCCATTCGATCCATCCAGCTTATGTTCTGGCCATCGCCATCAGCGCAGATGCGGC 339

Qy     17 GlyLeuIleArgLysLeuSerAspPheIleAspProGlnGluGlyTrpLysLeuAla 36
            :::|||:||||||||:::||||||||||||||||||||||||:::|||
Db    340 CAACTCCTTCGCCAGCTGGCAAAAGGCGGAGATCCGACCCGGAAGTTCACAAGGTGGC 399

Qy     37 ValAlaIleAlaLysLysPro---SerGlyAspAspArgTyrAsnGln 50
            :::||||||||:::||||||   :::||||||||||
Db    400 GCGACTTAACCACCCGGTTAATGGTGATGACGGATAATAATCAA 444
```

RESULT 15
ABL27970/c
ID   ABL27970
     ABL27970 standard; DNA; 3610 BP.
XX
AC   ABL27970;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 35383.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 35383; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL16176-ABL30511) and the encoded proteins
CC   (ABS57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3610 BP; 977 A; 756 C; 795 G; 1082 T; 0 other;

Alignment Scores:

Pred. No.:   24.7   Length: 3610
Score: 66.00   Matches: 19
Percent Similarity: 52.73%   Conservative: 10
Best Local Similarity: 34.55%   Mismatches: 20
Query Match: 20.43%   Indels: 6
DB: 23   Gaps: 2

US-10-001-254-26 (1-59) x ABL27970 (1-3610)

```
Qy      2 AsnLysProIleThrProSerThrTyrValArgCysLeuAsnVal------------- 16
          |||  |||||   || ||||||::||:||:|||||          ::::::
Db   2331 AATCCTCCCAATCGATCCATCCAGCTATGTTCTGGCCATCAGCCAGCAGATGCGGCGGGGC 2272

Qy     17 GlyLeuIleArgLysLeuSerAspPheIleAspProGlnGlyTrpLysLysLeuAla 36
          |||:|||:||||::||:|||||::||||::
Db   2271 CAACTCCTTCGCCCAGCTTGGCAAAAGCGGAGGATGCCCGCCGGAAGTTCACAAGGTGGCT 2212

Qy     37 ValAlaIleLysLysPro---SerGlyAspAspArgTyrAsnGln 50
          |||          :::::||||||||
Db   2211 GCGACTTTAACCACCCCGGTTAATGGTGATGACGATAATAATCAA 2167
```

Search completed: January 18, 2004, 01:32:43
Job time : 98.7617 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 01:16:08 ; Search time 15.3955 Seconds
(without alignments)
1691.512 Million cell updates/sec

Title: US-10-001-254-26
Perfect score: 323
Sequence: 1 MNKPITPSTYVRCLNVGLIR..........KKPSGDDRYNQFHIRCCSQN 59

Scoring table: BLOSUM62

		Xgapext	0.5
Xgapop	10.0		
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152425_19740/app_query.fasta_1.1109
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10001254 @CGN 1.1.95 @runat_16012004_152425_19740 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	290	89.8	833	4	US-09-166-350-10	Sequence 10, Appl
2	61	18.9	580073	4	US-08-545-528D-1	Sequence 1, Appli
3	59.5	18.4	3381	4	US-09-009-119-1	Sequence 1, Appli
4	59.5	18.4	3381	4	US-09-371-507-1	Sequence 1, Appli
5	59.5	18.4	3383	5	PCT-US95-09908-1	Sequence 1, Appli
6	58	18.0	1182	3	US-09-252-991A-4901	Sequence 4901, Ap
7	58	18.0	3482	3	US-09-111-085-3	Sequence 3, Appli
8	58	18.0	202001	4	US-09-734-674-3	Sequence 3, Appli
9	57.5	17.8	1794	5	US-08-123-934A-9	Sequence 9, Appli
10	57.5	17.8	1794	5	PCT-US94-11080-9	Sequence 9, Appli
11	57.5	17.8	3293	1	US-07-923-976-1	Sequence 1, Appli
12	57	17.6	2462	3	US-09-111-085-1	Sequence 1, Appli

13	56.5	17.5	936	3	US-09-461-697-402	Sequence 402, App
14	56.5	17.5	1017	3	US-09-461-697-396	Sequence 396, App
15	56.5	17.5	1128	3	US-09-461-697-394	Sequence 394, App
16	56.5	17.5	1260	3	US-09-461-697-388	Sequence 388, App
17	56.5	17.5	1467	3	US-09-461-697-374	Sequence 374, App
18	56.5	17.5	1473	3	US-09-461-697-372	Sequence 372, App
19	56.5	17.5	2702	3	US-09-461-697-359	Sequence 359, App
20	56	17.3	941	3	US-08-829-613-1	Sequence 1, Appli
21	56	17.3	1623	4	US-09-513-783A-33	Sequence 33, Appl
22	56	17.3	61663	4	US-09-453-702B-62	Sequence 62, Appl
23	55.5	17.2	1029	2	US-08-484-993B-21	Sequence 21, Appl
24	55.5	17.2	1029	2	US-08-484-158B-21	Sequence 21, Appl
25	55.5	17.2	1029	2	US-08-484-596A-21	Sequence 21, Appl
26	55.5	17.2	1029	2	US-08-480-150A-21	Sequence 21, Appl
27	55.5	17.2	1029	3	US-08-458-731-21	Sequence 21, Appl
28	55.5	17.2	1029	2	US-08-149-223A-21	Sequence 21, Appl
29	55.5	17.2	2238	2	US-08-674-887A-7	Sequence 7, Appli
30	55.5	17.2	2238	3	US-08-951-844-7	Sequence 7, Appli
31	55.5	17.2	2238	4	US-09-412-347-7	Sequence 7, Appli
32	55.5	17.2	2515	4	US-09-221-017B-745	Sequence 745, App
33	55.5	17.2	420	3	US-08-991-890-1	Sequence 1, Appli
34	55	17.0	438	4	US-08-979-847B-128	Sequence 128, App
35	55	17.0	438	4	US-08-979-847B-129	Sequence 129, App
36	55	17.0	438	4	US-08-979-847B-130	Sequence 130, App
37	55	17.0	438	4	US-08-979-847B-131	Sequence 131, App
38	55	17.0	480	4	US-08-991-890-3	Sequence 3, Appli
39	55	17.0	615	2	US-08-482-842B-1	Sequence 1, Appli
40	55	17.0	615	2	US-08-482-842B-3	Sequence 3, Appli
41	55	17.0	615	4	US-09-174-465D-1	Sequence 1, Appli
42	55	17.0	615	4	US-09-599-564A-1	Sequence 1, Appli
43	55	17.0	934	4	US-09-174-465D-4	Sequence 4, Appli
44	55	17.0	934	4	US-09-599-564A-4	Sequence 4, Appli
45	55	17.0	15144	3	US-08-458-434A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-166-350-10

Sequence 10, Application US/09166350A
Patent No. 6440663

; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Chen, Yao
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alex
; TITLE OF INVENTION: Renal Cancer Associated Antigens and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: L0461/7051
; CURRENT APPLICATION NUMBER: US/09/166,350A
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: US 09/166,350
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 10
; LENGTH: 833
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-166-350-10

Alignment Scores:
Pred. No.:
Score: 8.07e-36
     290.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 89.78%
DB:

Length: 833
Matches: 54
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-001-254-26 (1-59) x US-09-166-350-10 (1-833)

RESULT 2
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium

US-08-545-528D-1

Alignment Scores:
Pred. No.:
Score: 3.13e+04
Percent Similarity: 61.00   Matches: 12
Score: 46.94%   Conservative: 11
Best Local Similarity: 24.49%   Mismatches: 8
Query Match: 18.8%   Indels: 18
DB: 4   Gaps: 1
Length: 580073

US-10-001-254-26 (1-59) x US-08-545-528D-1 (1-580073)

Qy   1 MetAsnLysProIleThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArg 20

Qy   11 ValArgGlySerAsnValGlyLeuIleArgLysLeuSerAspPhe------------------ 25
       |||:|:| ||::|||||
Db   96165 ATTCATTGTTTAAACGTTTGCTCATGACAAGATTGAAACAGCGATTAGCAAAAATTAGGTA 96224

Qy   26 -------------------IleAspProGlnGluGlyTrp 32
       :::||||
Db   96225 ATGAAACATTGACAGTGCCTTCTTTTCCAAAAATGTTAAGTAGAACCCAAAAACGGGTGG 96284

Qy   33 LysLysLeuAlaValAlaIleIleLysLys 41
       |||
Db   96285 AAGCAATTAACTTTGACACTAGAAAAA 96311

RESULT 3
US-09-009-119-1/c
; Sequence 1, Application US/0909119
; Patent No. 6160206
; GENERAL INFORMATION:
; APPLICANT: SATO, Ryo
; APPLICANT: BOYNTON, John E.
; APPLICANT: GILLHAM, Nicholas W.
; APPLICANT: HARRIS, Elizabeth H.
; TITLE OF INVENTION: Porphyrin Accumulating-Type Herbicide Resistance Gene
; FILE REFERENCE: substitute sequence listing
; CURRENT APPLICATION NUMBER: US/09/009,119
; CURRENT FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3381
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii

US-09-009-119-1

Alignment Scores:
Pred. No.:
Score: 45.3   Length: 3381
Percent Similarity: 59.50   Matches: 17
Score: 46.77%   Conservative: 12
Best Local Similarity: 27.42%   Mismatches: 22
Query Match: 18.42%   Indels: 11
DB: 3   Gaps: 2

US-10-001-254-26 (1-59) x US-09-009-119-1 (1-3381)

Qy   6 ThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPhe 25
       :::|||: |||
Db   2156 AGCCCGGCTAGCCCTTTGCGGTGCTTGAGTGGGATGTGGGAGGTGCATTTGGGATATC 2097

Qy   26 IleAspProGlnGluGlyTrpLysLysLeu------------------ 35
       :::
Db   2096 ATGGACCGTGAAGTGGCGTGGGTGGGGGGGTGGGGGGGACGGGGCATGTCGGT 2037

RESULT 4
US-09-371-507-1/c
; Sequence 1, Application US/09371507
; Patent No. 6346656
; GENERAL INFORMATION:
; APPLICANT: SATO, Ryo
; APPLICANT: BOYNTON, John E.
; APPLICANT: GILLHAM, Nicholas W.
; APPLICANT: HARRIS, Elizabeth H.
; TITLE OF INVENTION: Porphyrin Accumulating-Type Herbicide Resistance Gene
; FILE REFERENCE: substitute sequence listing
; CURRENT APPLICATION NUMBER: US/09/371,507
; CURRENT FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 09/009,119
; PRIOR FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3381
; TYPE: DNA
; ORGANISM: Chlamydomonas reinhardtii

US-09-371-507-1

Alignment Scores:
Pred. No.:
Score: 45.3   Length: 3381
Percent Similarity: 59.50   Matches: 17
Score: 46.77%   Conservative: 12
Best Local Similarity: 27.42%   Mismatches: 22
Query Match: 18.42%   Indels: 11
DB: Gaps: 2

US-10-001-254-26 (1-59) x US-09-371-507-1 (1-3381)

Qy   6 ThrProSerThrThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPhe 25
       :::|||: |||
Db   2156 AGCCCGGCTAGCCCTTTGCGGTGCTTGAGTGGGATGTGGGAGGTGCATTTGGGATATC 2097

Qy   26 IleAspProGlnGluGlyTrpLysLysLeu------------------ 35
       :::
Db   2096 ATGGACCGTGAAGTGGCGTGGGTGGGGGGGTGGGGGGGACGGGGCATGTCGGT 2037

Qy   36 AlaValAlaIleIleLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCys 55
       |||||||| :|| ||| :|:
Db   2036 GCCTCGGCACAGCGTTGGCCTAGTGGCCAGCTCCCGCTGGATGGGCTTGCAA---GGGGTGC 1980

Qy   56 CysSer 57
       |||||
Db   1979 TGTTCA 1974

```
Qy    36 AlaValAlaIleLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCys 55
            ||||                                              ||:
Db  2036 GCCTCGGCACAGCGTTGGCCTAGTGGGCCAGTCCGCTGGATGGGCCTTGCAA---GGGTGC 1980

Qy    56 CysSer 57
            ||||||
Db  2037 GCCTCGGCACAGCGTTGGCCTAGTGGGCCAGTCCGCTGGATGGGCCTTGCAA---GGGTGC 1981

Qy    56 CysSer 57
            ||||||
Db  1979 TGTTCA 1974
```

RESULT 5
; Sequence 1, Application PC/TUS9509098
; PCT-US95-09098-1/c
; GENERAL INFORMATION:
; APPLICANT: Sato, Ryo
; APPLICANT: Boynton, John E.
; APPLICANT: Gillham, Nicholas W.
; APPLICANT: Harris, Elizabeth H.
; TITLE OF INVENTION: Porphyrin-Accumulating Type Herbicide
; TITLE OF INVENTION: Resistance Gene
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09098
; FILING DATE: 20-JUL-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 2185-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3383 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Chlamydomonas reinhardtii
; STRAIN: RS-3
PCT-US95-09098-1

Alignment Scores:
Pred. No.:
Score: 45.4
Percent Similarity: 59.50   Length: 3383
Best Local Similarity: 46.77%   Matches: 17
Query Match: 27.42%   Conservative: 12
DB: 18.42%   Mismatches: 22
                                            Indels: 11
                                            Gaps: 2

US-10-001-254-26 (1-59) x PCT-US95-09098-1 (1-3383)

```
Qy    6 ThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPhe 25
           ::||||:::       ||:::|||||||::
Db 2157 AGCCCGGCTAGCCCTTTGCGGTGCTTGAGTGGGATGGGGTTGCAGGGTGCCATTGGGATATC 2098

Qy    26 IleAspProGlnGluGlyThrTrpLysLysLeu--------------- 35
           ||:|||
```

```
Db 2097 ATGGACCGTGAAGTGGCGTGGGTAAGTGCCTGGCGTGGCGGGGACAGGGCCATGTCCGGT 2038

Qy    36 AlaValAlaIleLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCys 55
            ||||                                              ||:
Db 2037 GCCTCGGCACAGCGTTGGCCTAGTGGGCCAGTCCGCTGGATGGGCCTTGCAA---GGGTGC 1981

Qy    56 CysSer 57
            ||||||
Db 1980 TGTTCA 1975
```

RESULT 6
US-09-252-991A-4901/c
; Sequence 4901, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4901
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4901

Alignment Scores:
Pred. No.:
Score: 18.2
Percent Similarity: 58.00   Length: 1182
Best Local Similarity: 57.14%   Matches: 9
Query Match: 32.14%   Conservative: 7
DB: 17.96%   Mismatches: 12
                                            Indels: 0
                                            Gaps: 0

US-10-001-254-26 (1-59) x US-09-252-991A-4901 (1-1182)

```
Qy    29 GlnGluGlyThrTrpLysLysLeuAlaValAlaAlaIleLysLysProSerGlyAspAspArgTyr 48
            :::::|       :::|||:::|||
Db    89 AAACGGGCCCTGGGAGCCGTCTTGCGGTGCTTGACGGTCGGGGCGCAGGAGGACAACTGG 30

Qy    49 AsnGlnPheHisIleArgCys 56
            :::||||||
Db    29 TCCATGATCATCCGGAAGTGCTGC 6
```

RESULT 7
US-09-111-085-3/c
; Sequence 3, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: specific sequences
; FILE REFERENCE: 4238/75168
; CURRENT APPLICATION NUMBER: US/09/111,085
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: GB 9710154.7
; EARLIER FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 3
; SOFTWARE: Patentin Ver. 2.0
; LENGTH: 3482
; TYPE: DNA
; ORGANISM: Porcine retrovirus
US-09-111-085-3

Alignment Scores:

```
Pred. No.:
Score:                  80.8        Length:         3482
Percent Similarity:     58.00       Matches:        18
Best Local Similarity:  42.59%      Conservative:   5
Query Match:            33.33%      Mismatches:     23
DB:                     17.96%      Indels:         8
                                    Gaps:           3

US-10-001-254-26 (1-59) x US-09-111-085-3 (1-3482)

Qy    11 ValArgCysLeuAsnValGlyLeuIleArgGlyLeuSerAspPheIleAsp--------- 27
            ||||||
Db  1017 GTTAGGTAAGGAACCAGGGCCGATGGAGCCGAAGCTTAAGGGATTCTCAGTCTTTTCGGC 958

Qy    28 --------ProGlnGluGly---TrpLysLeuAlaValAlaIleLysLysProSerGly 44
            |||     |||:::    |||||||||||||
Db   957 TCTCCACCCGAGTCGGGAGTGGGCGCCCAGCTTAACGTGGGATGCATGGGATCCCAGGTGGA 898

Qy    45 AspAspArgTyrAsnGlnPheHisIleArgCysSerGln 58
            |||                 |||  |||
Db   897 GATTCC-------TTCGACTTTCACAGCCGTTGGTGTGGTCAA 862
```

RESULT 8
US-09-734-674-3/c
Sequence 3, Application US/09734674
Patent No. 6498022
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

```
Alignment Scores:
Pred. No.:
Score:                  2.15e+04    Length:         202001
Percent Similarity:     58.00       Matches:        13
Best Local Similarity:  62.96%      Conservative:   4
Query Match:            48.15%      Mismatches:     10
DB:                     17.96%      Indels:         0
                                    Gaps:           0

US-10-001-254-26 (1-59) x US-09-734-674-3 (1-202001)

Qy     29 GlnGluGlyTrpLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr 48
             |||
Db  91623 AAAGAGAGTAACAAAAAAATTAGCTGTACCTATAAAATACACTAAGGGGAAAAGAATATAC 91564

Qy     49 AsnGlnPheHisIleArgCys 55
             ||||||
Db  91563 CTGCAATTTCTTCTAAAGTGT 91543
```

RESULT 9
US-08-123-934A-9/c
Sequence 9, Application US/0812334A
Patent No. 6291206
; GENERAL INFORMATION:
; APPLICANT: WOZNEY, John
; APPLICANT: CELESTE, Anthony J.
; APPLICANT: THIES, R. Scott
; APPLICANT: YAMAJI, No. 6291206oru
; TITLE OF INVENTION: RECEPTOR PROTEINS

; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute Inc.- Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; ZIP: 02140
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/123,934A
; FILING DATE: 17-SEP-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: LAZAR, Steven R
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5203
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 876 1170
; TELEFAX: 617 876 5851
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1794 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 83..1591
US-08-123-934A-9

```
Alignment Scores:
Pred. No.:
Score:                  38.8        Length:         1794
Percent Similarity:     57.50       Matches:        18
Best Local Similarity:  43.86%      Conservative:   7
Query Match:            31.58%      Mismatches:     29
DB:                     17.80%      Indels:         3
                                    Gaps:           1

US-10-001-254-26 (1-59) x US-08-123-934A-9 (1-1794)

Qy      5 IleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArgGlyLeuSerAsp 24
            :::    ||||||      :::::  ||||||
Db   1587 TTGATCCTTCCTCGTTGGCTGAGTTGTGACAATGTTTTTAATTCGCAAAGCTGTCAGC 1528

Qy     25 PheIleAspPro---------GlnGluGlyTrpLysLeuAlaValAlaIleLysLys 41
            |||                 |||
Db   1527 CTTGCTGCTCCATTCGCATACCAGCATTCTCTCATTAATTTAGCCATCACTCTCAAGGCC 1468

Qy     42 ProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysSerGln 58
            |||                 :::
Db   1467 TCACAGCTCTCGCCATCTGTTTGGAATATTGGCCTTAACTTCTGTTCGCAA 1417
```

RESULT 10
PCT-US94-10080-9/c
Sequence 9, Application PC/TUS9410080
; GENERAL INFORMATION:
; APPLICANT: GENETICS INSTITUTE, INC.
; TITLE OF INVENTION: RECEPTOR PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute Inc.- Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA

; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10080
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,934
; FILING DATE: 17-SEP-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: LAZAR, Steven R
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5203-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8260
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1794 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: W-120
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 83..1591
PCT-US94-10080-9

Alignment Scores:
Pred. No.:
Score: 38.8   Length: 1794
Percent Similarity: 57.50   Matches: 18
Best Local Similarity: 43.86%   Conservative: 7
Query Match: 31.58%   Mismatches: 29
DB: 17.80%   Indels: 3
                                       Gaps: 1

US-10-001-254-26 (1-59) x PCT-US94-10080-9 (1-1794)

Qy   5 IleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAsp 24
            ::::||| ::::::|
Db   1587 TTGATGCCTTCCTGTGTGGCTGAGTTGTGACAAGTTTTTTTAATTCGCAAAGCTGTCAGC 1528

Qy   25 PheIleAspPro------GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysLys 41
            ||||| |||||||::::   ::|||
Db   1527 CTTGCTGCTTCCATTGGCATACCAGCAATTCTCTCATTAATTTTTAGCCATCACTCAAGGCC 1468

Qy   42 ProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysCysSerGln   58
            ::: ||| :::
Db   1467 TCACAGCTCTCGCCATCTGTTTGGAATATTTGGCCTTAACTTCCTGTTCGGCAA   1417

RESULT 11
US-07-923-976-1/c
; Sequence 1, Application US/0792397 6
; Patent No. 5574136
; GENERAL INFORMATION:
; APPLICANT: Nagata, Shigekazu
; APPLICANT: Fukunaga, Rikiro
; TITLE OF INVENTION: DNA Encoding Granulocyte-
; TITLE OF INVENTION: Colony-Stimulating Factor Receptor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones, Tullar & Cooper, P.C.
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; ZIP: 2202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/923,976
; FILING DATE: 19920922
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 74539/1990
; FILING DATE: 23-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 176629/1990
; FILING DATE: 03-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/00375
; FILING DATE: 22-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hellwege, James W.
; REGISTRATION NUMBER: 28,808
; REFERENCE/DOCKET NUMBER: 514853
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-415-1500
; TELEFAX: 703-415-1508
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3293 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 180..2690
US-07-923-976-1

Alignment Scores:
Pred. No.:
Score: 89.6   Length: 3293
Percent Similarity: 57.50   Matches: 19
Best Local Similarity: 35.82%   Conservative: 5
Query Match: 28.36%   Mismatches: 14
DB: 17.80%   Indels: 29
                                       Gaps: 3

US-10-001-254-26 (1-59) x US-07-923-976-1 (1-3293)

Qy   12 ArgCysLeuAsnValGlyLeuIleArgLys------------------------------ 21
            ::: ||||| ::|:::
Db   507 CGCTGTCTTCCCATGGCACTAAGCACAGAAGAGGAAGGCCTGGGTGTAGTTCAAGTGAGGCA 448

Qy   22 ------------------------------LeuSerAspPheIleAspProGlnGluGly 31
                                           :::||||||| |||||
Db   447 GAGTGATGAGGGACTCTTGGGTCCCATCAGCGACAGATGATGCTGTCTGTCCCAGGTTGGA 388

Qy   32 TrpLysLysLeuAlaValAlaIleLysLys---ProSerGlyAspAspArgTyrAsnGln 50
            ||| ||||||:::|||::::: ||| |||
Db   387 TGGGCTCATCTTGCAGTCTCCCATAAGATCTTTGCCTGTTGGTCCTGTTGGTCCA   343

Qy   51 PheHisIleArgCysCysSer   57
               ||||| |||||||
Db   342 ----------GTTTGCTGCAGT   331

RESULT 12
US-09-111-085-1/c
; Sequence 1, Application US/09111085
; Patent No. 6100034
; GENERAL INFORMATION:
; APPLICANT: Stoye, Jonathan P
; APPLICANT: Weiss, Robin A
; TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope
; TITLE OF INVENTION: Specific sequences
; FILE REFERENCE: 4238/75168

```
;  CURRENT APPLICATION NUMBER: US/09/111,085
;  CURRENT FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: GB 9710154.7
;  EARLIER FILING DATE: 1997-05-16
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.0
;
;  SEQ ID NO 1
;  LENGTH: 2462
;  TYPE: DNA
;  ORGANISM: Porcine retrovirus
US-09-111-085-1

Qy   DB:

Alignment Scores:
Pred. No.:
Score:                    71.8            Length:        2462
Percent Similarity:       57.00           Matches:       19
Best Local Similarity:    39.34%          Conservative:  5
Query Match:              31.15%          Mismatches:    29
                          17.65%          Indels:        8
                                          Gaps:          3

US-10-001-254-26 (1-59) x US-09-111-085-1 (1-2462)

Qy   4  ProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSer  23
           ||| :::
Db  338  CCAATAACTTGAGGAGTTAATTGACAGAGTAAGGAACCACGGCGATGGAGGCGGAACCTTAAG  279

Qy  24  AspPheIleAsp----------------ProGlnGluGly---TrpLysLysLeuAlaVal  37
                                      ||||||          :::|||
Db  278  GGGGATTTCAGTCTTTTCCGGCCTTTCCACCCCCGAATCGGGAGGGTGGCGCCCCGGCTTAACGTG  219

Qy  38  AlaIleLysLysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysCysSer  57
                                      |||           |||||||
Db  218  GGGATGCATGGAATCCAGGTGGGGGAATTCC------TTCGACTTTCACAGCCCTTGGTGTGGT  165

Qy  58  Gln 58
         |||
Db  164  CAA 162
```

```
RESULT 13
US-09-461-697-402
Sequence 402, Application US/09461697
;  Patent No. 6277974
;  GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
;  FILE REFERENCE: 1000-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
;
SEQ ID NO 402
;  LENGTH: 936
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-461-697-402

Db

Alignment Scores:
Pred. No.:
Score:                    22.7            Length:        936
Percent Similarity:       56.50           Matches:       21
Best Local Similarity:    31.46%          Conservative:  7
Query Match:              23.60%          Mismatches:    26
                          17.49%          Indels:        35
                                          Gaps:          2

US-10-001-254-26 (1-59) x US-09-461-697-402 (1-936)
```

```
RESULT 14
US-09-461-697-396
Sequence 396, Application US/09461697
;  Patent No. 6277974
;  GENERAL INFORMATION:
;  APPLICANT: COGENT NEUROSCIENCE, Inc.
;  APPLICANT: Lo, Donald C.
;  APPLICANT: Barney, Shawn
;  APPLICANT: Thomas, Mary Beth
;  APPLICANT: Portbury, Stuart D.
;  APPLICANT: Puranam, Kasturi
;  APPLICANT: Katz, Lawrence C.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
;  TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
;  TITLE OF INVENTION: CELL DEATH
;  FILE REFERENCE: 10001-005-999
;  CURRENT APPLICATION NUMBER: US/09/461,697
;  CURRENT FILING DATE: 1999-12-14
;  NUMBER OF SEQ ID NOS: 466
;  SOFTWARE: FastSEQ for Windows Version 4.0
;
SEQ ID NO 396
;  LENGTH: 1017
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-461-697-396

DB:

Alignment Scores:
Pred. No.:
Score:                    25.4            Length:        1017
Percent Similarity:       56.50           Matches:       21
Best Local Similarity:    31.46%          Conservative:  7
Query Match:              23.60%          Mismatches:    26
                          17.49%          Indels:        35
                                          Gaps:          2

US-10-001-254-26 (1-59) x US-09-461-697-396 (1-1017)

Qy   1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsn------------  15
           ||||||||||||
Db   1  ATGAATAAAGCTCTACTACCTTCTCTTTGCGGTGCTATGTGGAAATGACCATGTTAATCTA  60

Qy  15  ------------------------------------------------------------  15
                                                       :::
Db  61  CCCATCATGTGAGACAATTCTTAAGTAAAAGCCGTTCTTCCTCTTGGAGCTACCAGTTCTAA  120

Qy  16  ------------------------ValGlyLeuIleArgLysLeuSerAspPheIleAspPro  28
                                    :::||||||||         ||||
Db 121  GGGGAGGAGAGACCACCACGGAATCCTGGGACTTCTGAATTGGTTGTCTCATTTTGCCAACCCT  180

Qy  29  GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr  48
           |||        :::   |||                    ||||||||
Db 181  ACCGGAAGCACTAGATAATGTTCTGAAATAACCTCCCAAAA-------AAGGATCGAGAA  231

Qy  49  AsnGlnPheHisIleArgCysCysSer  57
           |||        :::
Db 232  AATGTTAAGGAACTTCTCGCTGTTCC  258
```

```
Qy   1  MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsn------------  15
           |||||||||||||
Db   1  ATGAATAAAGCTCTACTACCCTTCTTGCGGTGCTATGTGGAAATGACCATGTTAATCTA  60

Qy  15  ------------------------------------------------------------  15
Db  82  ATGAATAAAGCTCTACTACCCTTCTTGCGGTGCTATGTGGAAAATGACCATGTTAATCTA  141

Qy  15  ------------------------ValGlyLeuIleArgLysLeuSerAspPheIleAspPro  28
                                    :::||||||||         ||||
Db 142  CCCATCATGAGAGACAATTCTTAAGTAAAAGCCGTTCTTCCTCTTGGAGCTACCAGTTCTAA  201

Qy  29  GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr  48
           |||        :::                          ||||||||
Db 202  GGGGAGGAGAGACCACCACGAATCCTGGGACTTCTGAATTGGTTGTCTCATTTTGCCAACCCT  261

Qy  29  GlnGluGlyTrpLysLysLeuAlaValAlaIleLysLysProSerGlyAspAspArgTyr  48
           |||
Db 262  ACCGGAAGCACTAGATAATGTTCTGAAATAACCTCCCAAAA-------AAGGATCGAGAA  312
```

```
Qy        49 AsnGlnPheHisIleArgCysCysSer 57
             |||            |||||||||||
Db       313 AATGTTAAGGAACTTCTGCTGTTCC 339
```

RESULT 15
US-09-461-697-394

; Sequence 394, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 394
; LENGTH: 1128
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-394

Alignment Scores:
Pred. No.:
Score: 29.3   Length: 1128
Percent Similarity: 56.50   Matches: 21
Best Local Similarity: 31.46%   Conservative: 7
Query Match: 23.60%   Mismatches: 26
DB: 17.49%   Indels: 35
                            3   Gaps: 2

US-10-001-254-26 (1-59) x US-09-461-697-394 (1-1128)

```
Qy         1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsn------------------ 15
             ||||||||||||                 :::                 |||
Db       193 ATGAATAAAGCTCTACTACCTCTCTTGCGTGCTATGTGAAATGACCACCGTTAATCTA 252

Qy        15 ------------------------------------------------------------- 15
                                                                 |||
Db       253 CCCATCATGGAGACATTCTTAAGTAAAGGGCGTCTTCCTCTTGGAGCTACCAGTTCTAAA 312

Qy        16 ------------------ValGlyLeuIleArgLysLeuSerAspPheIleAspPro 28
                               ::|||||||||:::             :::::||
Db       313 GGGAGGAGGACACCACCGAAATCCTGGGACTTCTGAATTGGTTGTCTCATTTTGCCAACCT 372

Qy        29 GlnGluGlyTrpLysLeuAlaValAlaIleLeuLysGlyProSerGlyAspAspArgTyr 48
             |||              :::               |||
Db       373 ACCGAAGCACTAGATAATGTTCTGAAATACCTCCCAAAA-----AAGGATCGAGA 423

Qy        49 AsnGlnPheHisIleArgCysCysSer 57
             |||            |||||||||||
Db       424 AATGTTAAGGAACTTCTGCTGTTCC 450
```

Search completed: January 18, 2004, 06:11:39
Job time : 73.3955 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 04:27:15 ; Search time 88.2609 Seconds
(without alignments)
2356.207 Million cell updates/sec

Title: US-10-001-254-26
Perfect score: 323
Sequence: 1 MNKPIPSTYVRCLNVGLIR...........KKPSGDRYNQFHIRCCSQN 59

Scoring table: BLOSUM62
                Xgapop 10.0 / Xgapext 0.5
                Ygapop 10.0 / Ygapext 0.5
                Fgapop 6.0 / Fgapext 7.0
                Delop 6.0 / Delext 7.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10001254/runat_16012004_152426_19815/app_query.fasta_1.1109
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10001254 @CGN 1.1_534 @runat_16012004_152426_19815
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
            1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
            2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
            3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
            4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
            5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
            6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
            7: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
            8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
            9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
           10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
           11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
           12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
           13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
           14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
           15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
           16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
           17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
           18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Result
No.   Score Query   Match Length DB   ID   Description
---   ----- -----   ------------ --- --   -----------
                       %

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

1:	323	100.0	211	15	US-10-001-254-25	Sequence 25, Appli
2:	290	89.8	1383	11	US-09-759-595-2	Sequence 13929, A
3:	290	89.8	1383	15	US-10-001-254-15	Sequence 15, Appl
4:	290	89.8	2817	11	US-09-966-451-3	Sequence 3, Appli
5:	290	89.8	2817	15	US-10-001-254-27	Sequence 27, Appl
6:	290	89.8	31000	11	US-09-966-451-10	Sequence 10, Appl
7:	277	85.8	501	9	US-09-833-790-149	Sequence 149, App
8:	271	83.9	1542	11	US-09-759-595-4	Sequence 4, Appli
9:	247	76.5	294	15	US-10-001-254-5	Sequence 5, Appli
10:	73	22.6	497	13	US-10-027-632-50723	Sequence 50723, A
11:	73	22.6	497	13	US-10-027-632-50724	Sequence 50724, A
12:	73	22.6	497	14	US-10-027-632-50723	Sequence 50723, A
13:	73	22.6	497	14	US-10-027-632-50724	Sequence 50724, A
14:	73	22.6	499	13	US-10-027-632-72567	Sequence 72567, A
15:	73	22.6	499	13	US-10-027-632-72568	Sequence 72568, A
16:	73	22.6	499	14	US-10-027-632-72567	Sequence 72567, A
17:	73	22.6	499	14	US-10-027-632-72568	Sequence 72568, A
18:	68	21.1	5170	11	US-09-374-046A-131	Sequence 131, App
19:	65.5	20.3	4151	9	US-09-764-853-230	Sequence 230, App
20:	64	19.8	2388	13	US-09-870-133-3	Sequence 3, Appli
21:	64	19.8	2388	15	US-10-160-501-6	Sequence 6, Appli
22:	64	19.8	2583	10	US-09-976-674-6	Sequence 6, Appli
23:	64	19.8	3238	15	US-09-870-133-1	Sequence 1, Appli
24:	64	19.8	3238	15	US-10-369-493-29885	Sequence 29885, A
25:	63	19.5	665	13	US-10-027-632-156803	Sequence 156803,
26:	63	19.5	665	13	US-10-027-632-156804	Sequence 156804,
27:	63	19.5	665	13	US-10-027-632-156805	Sequence 156805,
28:	63	19.5	665	14	US-10-027-632-156803	Sequence 156803,
29:	63	19.5	665	14	US-10-027-632-156804	Sequence 156804,
30:	63	19.5	665	14	US-10-027-632-156805	Sequence 156805,
31:	62	19.2	260	12	US-10-369-493-30409	Sequence 30409, A
32:	62	19.2	1489	15	US-10-369-493-29885	Sequence 29885, A
33:	61.5	19.0	304	9	US-09-864-761-18690	Sequence 18690, A
34:	61.5	19.0	472	9	US-09-864-761-13929	Sequence 13929, A
35:	61	19.0	491	13	US-09-918-995-30747	Sequence 30747, A
36:	61	18.9	6109	13	US-09-795-061-1	Sequence 1, Appli
37:	61	18.9	13894	13	US-09-960-858-15	Sequence 15, Appl
38:	61	18.9	13894	13	US-09-960-870-15	Sequence 15, Appl
39:	61	18.9	580073	13	US-10-205-220-1	Sequence 1, Appli
40:	60.5	18.7	271990	13	US-10-345-072-87	Sequence 87, Appl
41:	60.5	18.7	271990	15	US-10-195-144-87	Sequence 87, Appl
42:	60.5	18.4	2361	12	US-10-108-260A-1536	Sequence 1536, Ap
43:	59.5	18.3	34	11	US-09-759-595-5	Sequence 5, Appli
44:	59	18.3	498	11	US-09-911-904-163	Sequence 163, App
45:	58.5	18.1	976	10	US-09-974-300-1559	Sequence 1559, Ap

ALIGNMENTS

RESULT 1
US-10-001-254-25

Sequence 25, Application US/10001254
; Sequence 25, Application US/10001254
; Publication No. US20030049702A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sug Hyung
; APPLICANT: Roth, Wilfred
; APPLICANT: Stenner-Liewen, Frank
; TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins
; FILE REFERENCE: P-LJ 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 211

```
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(177)
US-10-001-254-25

Alignment Scores:
Pred. No.:
Score:                          4.5e-43
Percent Similarity:             323.00
Best Local Similarity:          100.00%
Query Match:                    100.00%
DB:                             100.00%

US-10-001-254-26  (1-59) x US-10-001-254-25 (1-211)

Qy      1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAACAAACCCATAACACCATCAACACATATGTCGGCTCGCTTCAATGTTGGACTAATTAGG 60

Qy     21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleIleLys 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTATTAAA 120

Qy     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArgCysCysSerGlnAsn 59
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGAGCTGTGTTCCCAAAAC 177
```

```
RESULT 2
US-09-759-595-2
;   Sequence 2, Application US/09759595
;   Publication No. US20030059916A1
;   GENERAL INFORMATION:
;   APPLICANT: Wesche, Holger
;   APPLICANT: Li, Shyun
;   APPLICANT: Tularik Inc.
;   TITLE OF INVENTION: IRAK-4: Compositions and Methods of Use
;   FILE REFERENCE: 018781-0039100US
;   CURRENT APPLICATION NUMBER: US/09/759,595
;   CURRENT FILING DATE: 2001-01-13
;   PRIOR APPLICATION NUMBER: US 60/176,395
;   PRIOR FILING DATE: 2000-01-13
;   NUMBER OF SEQ ID NOS: 7
;   SOFTWARE: Patentin Ver. 2.1
;
;   SEQ ID NO 2
;   LENGTH: 1383
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION:
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1383)
;   OTHER INFORMATION: human IRAK-4
US-09-759-595-2

Alignment Scores:
Pred. No.:
Score:                          1.3e-36
Percent Similarity:             290.00
Best Local Similarity:          100.00%
Query Match:                    89.78%
DB:                             11

US-10-001-254-26  (1-59) x US-09-759-595-2 (1-1383)

Qy      1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAACAAACCCATAACACCATCAACACATATGTCGGCTCGCTGCCTCAATGTTGGACTAATTAGG 60

Qy     21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleIleLys 40
```

```
Db     61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTATTAAA 120
Qy     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
          ||||||||||||||||||||||||||||||||||||||||||
Db    121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 162
```

```
RESULT 3
US-10-001-254-15
;   Sequence 15, Application US/10001254
;   Publication No. US20030049702A1
;   GENERAL INFORMATION:
;   APPLICANT: Reed, John C.
;   APPLICANT: Godzik, Adam
;   APPLICANT: Pawlowski, Krzysztof
;   APPLICANT: Fiorentino, Loredana
;   APPLICANT: Lee, Sug Hyung
;   APPLICANT: Roth, Wilfred
;   APPLICANT: Stenner-Liewen, Frank
;   TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins
;   FILE REFERENCE: P-LJ 5037
;   CURRENT APPLICATION NUMBER: US/10/001,254
;   CURRENT FILING DATE: 2001-11-15
;   PRIOR APPLICATION NUMBER: 60/301,889
;   PRIOR FILING DATE: 2001-06-29
;   PRIOR APPLICATION NUMBER: 09/715,893
;   PRIOR FILING DATE: 2000-11-17
;   NUMBER OF SEQ ID NOS: 62
;   SOFTWARE: FastSEQ for Windows Version 4.0
;
;   SEQ ID NO 15
;   LENGTH: 1383
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1380)
US-10-001-254-15

Alignment Scores:
Pred. No.:
Score:                          1.3e-36
Percent Similarity:             290.00
Best Local Similarity:          100.00%
Query Match:                    89.78%
DB:                             15

US-10-001-254-26  (1-59) x US-10-001-254-15 (1-1383)

Qy      1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGAACAAACCCATAACACCATCAACACATATGTCGGCTCGCTGCCTCAATGTTGGACTAATTAGG 60

Qy     21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleIleLys 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGCTGTCAGATTTTATTGATCCTCAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTATTAAA 120

Qy     41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
          ||||||||||||||||||||||||||||||||||||||||||
Db    121 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 162
```

```
RESULT 4
US-09-966-451-3
;   Sequence 3, Application US/09966451
;   Publication No. US20030087856A1
;   GENERAL INFORMATION:
;   APPLICANT: C. Frank Bennett
;   APPLICANT: Susan M. Freier
;   FILE REFERENCE: RTS-0324
;   TITLE OF INVENTION: ANTISENSE MODULATION OF IL-1 RECEPTOR-ASSOCIATED KINASE-4 EXPRESS
;   CURRENT APPLICATION NUMBER: US/09/966,451
;   CURRENT FILING DATE: 2001-09-28
;   NUMBER OF SEQ ID NOS: 88
;   SEQ ID NO 3
```

```
;  LENGTH: 2817
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (50)...(1432)
   US-09-966-451-3

   Alignment Scores:
   Pred. No.:
   Score:                    3.43e-36
   Percent Similarity:       100.00
   Best Local Similarity:    100.00%
   Query Match:              89.78%
   DB:                       11

                             Length:         2817
                             Matches:        54
                             Conservative:   0
                             Mismatches:     0
                             Indels:         0
                             Gaps:           0

US-10-001-254-26 (1-59) x US-09-966-451-3 (1-2817)

Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    50 ATGAACAAACCCATAACACCATCAACACATATGTGCGGCTGCCTCAATGTTGGACTAATTAGG 109

Qy    21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleAlaIleLys 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   110 AAGCTGTCAGATTTTATTGATCCTCAAGAACGGATGGAAGAAGTTAGCTGTAGCTATTAAA 169

Qy    41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
           ||||||||||||||||||||||||||||||||||||||||||
Db   170 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 211

RESULT 5
US-10-001-254-27
;  Sequence 27, Application US/10001254
;  Publication No. US20030049702A1
;  GENERAL INFORMATION:
;  APPLICANT: Reed, John C.
;  APPLICANT: Godzik, Adam
;  APPLICANT: Pawlowski, Krzysztof
;  APPLICANT: Fiorentino, Loredana
;  APPLICANT: Lee, Sug Hyung
;  APPLICANT: Roth, Wilfred
;  APPLICANT: Stenner-Liewen, Frank
;  TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins
;  FILE REFERENCE: P-LJ 5037
;  CURRENT APPLICATION NUMBER: US/10/001,254
;  CURRENT FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: 60/301,889
;  PRIOR FILING DATE: 2001-06-29
;  PRIOR APPLICATION NUMBER: 09/715,893
;  PRIOR FILING DATE: 2000-11-17
;  NUMBER OF SEQ ID NOS: 62
;  SOFTWARE: FastSEQ for Windows Version 4.0
   SEQ ID NO 27
;  LENGTH: 2817
;  TYPE: DNA
;  ORGANISM: Homo sapien
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (50)...(1429)
   US-10-001-254-27

   Alignment Scores:
   Pred. No.:
   Score:                    3.43e-36
   Percent Similarity:       290.00
   Best Local Similarity:    100.00%
   Query Match:              89.78%
   DB:                       15

                             Length:         2817
                             Matches:        54
                             Conservative:   0
                             Mismatches:     0
                             Indels:         0
                             Gaps:           0

Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    50 ATGAACAAACCCATAACACCATCAACACATATGTGCGGCTGCCTCAATGTTGGACTAATTAGG 109

Qy    21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleAlaIleLys 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   110 AAGCTGTCAGATTTTATTGATCCTCAAGAACGGATGGAAGAAGTTAGCTGTAGCTATTAAA 169

Qy    41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
           ||||||||||||||||||||||||||||||||||||||||||
Db   170 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 211

RESULT 6
US-09-966-451-10
;  Sequence 10, Application US/09833790
;  Publication No. US20030087856A1
;  GENERAL INFORMATION:
;  APPLICANT: C. Frank Bennett
;  APPLICANT: Susan M. Freier
;  TITLE OF INVENTION: ANTISENSE MODULATION OF IL-1 RECEPTOR-ASSOCIATED KINASE-4 EXPRES
;  FILE REFERENCE: RTS-0324
;  CURRENT APPLICATION NUMBER: US/09/966,451
;  CURRENT FILING DATE: 2001-09-28
;  NUMBER OF SEQ ID NOS: 88
   SEQ ID NO 10
;  LENGTH: 31000
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
   US-09-966-451-10

   Alignment Scores:
   Pred. No.:
   Score:                    9.03e-35
   Percent Similarity:       290.00
   Best Local Similarity:    100.00%
   Query Match:              89.78%
   DB:                       11

                             Length:         31000
                             Matches:        54
                             Conservative:   0
                             Mismatches:     0
                             Indels:         0
                             Gaps:           0

US-10-001-254-26 (1-59) x US-09-966-451-10 (1-31000)

Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  9860 ATGAACAAACCCATAACACCATCAACACATATGTGCGGCTGCCTCAATGTTGGACTAATTAGG 9919

Qy    21 LysLeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleAlaIleLys 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  9920 AAGCTGTCAGATTTTATTGATCCTCAAGAACGGATGGAAGAAGTTAGCTGTAGCTATTAAA 9979

Qy    41 LysProSerGlyAspAspArgTyrAsnGlnPheHisIleArg 54
           ||||||||||||||||||||||||||||||||||||||||||
Db  9980 AAACCATCTGGTGATGATAGATACAATCAGTTTCACATAAGG 10021

RESULT 7
US-09-833-790-149
;  Sequence 149, Application US/09833790
;  Patent No. US20020068288A1
;  GENERAL INFORMATION:
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Wang, Tongtong
;  APPLICANT: Secrist, Heather
;  APPLICANT: Mohamath, Raodoh
;  APPLICANT: Indrias, Carol Y.
;  APPLICANT: Fan, Liqun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.512
;  CURRENT APPLICATION NUMBER: US/09/833,790
;  CURRENT FILING DATE: 2001-04-11
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: FastSEQ for Windows Version 4.0
   SEQ ID NO 149
;  LENGTH: 501
;  TYPE: DNA
;  ORGANISM: Homo sapien
```

```
US-10-001-254-26 (1-59) x US-10-001-254-27 (1-2817)

Qy     1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(501)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-790-149

Alignment Scores:
Pred. No.:
Score: 4.17e-35
Percent Similarity: 277.00   Length: 501
Best Local Similarity: 98.18%   Matches: 54
Query Match: 98.18%   Conservative: 0
DB: 85.76%   Mismatches: 0
                           9   Indels: 1
                                           Gaps: 0

US-10-001-254-26 (1-59) x US-09-833-790-149 (1-501)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ATGAACAAACCCATFAACACCATCAACACATATGTCCGCTGCCTCCAATGTTGGACTTATTAGG 80

Qy   21 LysLeuSerAspPheIleAspProGlnGlyGlyTyrTrpLysLysLeuAlaValAlaIleLeuAlaIleLeuLys 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 AAGCTGTCAGATTTTATTGATCCTCAAGGAGGATGGAAGAAGTTAGCTGTAGCTATTTAAA 140

Qy   41 LysProSerGlyAspAspArgTyrAsnGln PheHisIleArg 54
          ||||||||||||||||||||||||||||||||||
Db   141 AAACCATCTGGTGATGATAGATACAATCAAGTTTCACATAAGG 183

RESULT 8
US-09-759-595-4
; Sequence 4, Application US/09759595
; Publication No. US20030059916A1
; GENERAL INFORMATION:
; APPLICANT: Wesche, Holger
; APPLICANT: Li, Shyun
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: IRAK-4: Compositions and Methods of Use
; FILE REFERENCE: 018781-00391OUS
; CURRENT APPLICATION NUMBER: US/09/759,595
; CURRENT FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 60/176,395
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 4
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: murine IL-1 receptor-associated kinase 4 (IRAK-4)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (163)..(1542)
; OTHER INFORMATION: murine IRAK-4
US-09-759-595-4

Alignment Scores:
Pred. No.:
Score: 1.81e-33
Percent Similarity: 271.00   Length: 1542
Best Local Similarity: 98.15%   Matches: 49
Query Match: 90.74%   Conservative: 4
DB: 83.90%   Mismatches: 1
                           11   Indels: 0
                                           Gaps: 0

US-10-001-254-26 (1-59) x US-09-759-595-4 (1-1542)

Qy   1 MetAsnLysProIleThrProSerThrTyrValArgCysLeuAsnValGlyLeuIleArg 20
          ||||||||||||||||||||||||||||||||||||||::||||
Db   163 ATGAACAAGCCGTTGACACCATCGACATACGACACCTTAATGTGGGGATCCTTAGG 222

Qy   21 LysLeuSerAspPheIleAspProGlnGlyGlyTyrTrpLysLysLeuAlaValAlaIleLeuAlaIleLeuLys 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||

Db   223 AAGCTGTCGGATTTTATTGATCCTCAAGGAGGGTGGAAGAAAATTAGCAGTAGCTATCAAA 282

Qy   41 LysProSerGlyAspAspArgTyrAsnGln PheHisIleArg 54
          |||||||||||||||||||||||||||||||||||||||
Db   283 AAGCCGTCCGGCGACGACAGATACAAATCAGTTCCATATAAGG 324

RESULT 9
US-10-001-254-5
; Sequence 5, Application US/10001254
; Publication No. US20030049702A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sug Hyung
; APPLICANT: Roth, Wilfred
; APPLICANT: Stenner-Liewen, Frank
; TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins
; FILE REFERENCE: P-LJ 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 5
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(294)
US-10-001-254-5

Alignment Scores:
Pred. No.:
Score: 1.47e-30
Percent Similarity: 247.00   Length: 294
Best Local Similarity: 100.00%   Matches: 46
Query Match: 100.00%   Conservative: 0
DB: 76.47%   Mismatches: 0
                           15   Indels: 0
                                           Gaps: 0

US-10-001-254-26 (1-59) x US-10-001-254-5 (1-294)

Qy   9 ThrTyrValArgCysLeuAsnValGlyLeuIleArgLysLeuSerAspPheIleAspPro 28
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACATATGTCCGCTGCCTCCAATGTTGGACTTATTAGGAAGCTGTCAGATTTTATTGATCCT 60

Qy   29 GlnGlyGlyTyrTrpLysLysLeuAlaValAlaIleLeuAlaIleLeuLysLysProSerGlyAspAspArgTyr 48
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CAAGGAGGATGGAAGAAGTTAGCTGTAGCTATTTAAAAAAACCATCTGGTGATGATAGATAC 120

Qy   49 AsnGlnPheHisIleArg 54
          ||||||||||||||
Db   121 AATCAGTTTCACATAAGG 138

RESULT 10
US-10-027-632-50723/c
; Sequence 50723, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676

```
          PRIOR FILING DATE: 2000-04-20
          PRIOR APPLICATION NUMBER: US 60/193,483
          PRIOR FILING DATE: 1999-11-23
          PRIOR APPLICATION NUMBER: US 60/156,358
          PRIOR FILING DATE: 1999-09-28
          PRIOR APPLICATION NUMBER: US 60/146,002
          PRIOR FILING DATE: 1999-08-09
          NUMBER OF SEQ ID NOS: 325720
          SOFTWARE: FastSEQ for Windows Version 4.0
   ;
   ;  SEQ ID NO 50723
   ;  LENGTH: 497
   ;  TYPE: DNA
   ;  ORGANISM: Human
   US-10-027-632-50723

   Alignment Scores:
   Pred. No.:               0.0477      Length:          497
   Score:                   73.00       Matches:         12
   Percent Similarity:      69.5%       Conservative:    4
   Best Local Similarity:   52.1%       Mismatches:      7
   Query Match:             22.60%      Indels:          0
   DB:                      13          Gaps:            0

   US-10-001-254-26 (1-59) x US-10-027-632-50723 (1-497)

   Qy    22 LeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLysLys 41
             |||   ||::|       |||  ||||||||
   Db   304 CTTCTAGATTTTGTGTCCCCTKACAAGGGTTGGAAGAGTCTGGCTCGAGCCCTGGCCTCT 245

   Qy    42 ProSerGly 44
             |||:||:|||
   Db   244 CCAGCTGGC 236

   RESULT 11
   US-10-027-632-50724/c
   Sequence 50724, Application US/10027632
   Publication No. US20030204075A9
   GENERAL INFORMATION:
   APPLICANT: Wang, David G.
   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
   TITLE OF INVENTION: Polymorphisms in the Human Genome
   FILE REFERENCE: 108827.129
   CURRENT APPLICATION NUMBER: US/10/027,632
   CURRENT FILING DATE: 2002-04-30
          PRIOR APPLICATION NUMBER: US 60/218,006
          PRIOR FILING DATE: 2000-07-12
          PRIOR APPLICATION NUMBER: US 60/198,676
          PRIOR FILING DATE: 2000-04-20
          PRIOR APPLICATION NUMBER: US 60/193,483
          PRIOR FILING DATE: 1999-11-23
          PRIOR APPLICATION NUMBER: US 60/167,363
          PRIOR FILING DATE: 2000-03-29
          PRIOR APPLICATION NUMBER: US 60/185,218
          PRIOR FILING DATE: 2000-02-24
          PRIOR APPLICATION NUMBER: US 60/156,358
          PRIOR FILING DATE: 1999-09-28
          PRIOR APPLICATION NUMBER: US 60/146,002
          PRIOR FILING DATE: 1999-08-09
          NUMBER OF SEQ ID NOS: 325720
          SOFTWARE: FastSEQ for Windows Version 4.0
   ;
   ;  SEQ ID NO 50724
   ;  LENGTH: 497
   ;  TYPE: DNA
   ;  ORGANISM: Human
   US-10-027-632-50724

   Alignment Scores:
   Score:                   73.00       Length:          497
   Pred. No.:               0.0477      Matches:         12
```

```
   Percent Similarity:      69.5%       Conservative:    4
   Best Local Similarity:   52.1%       Mismatches:      7
   Query Match:             22.60%      Indels:          0
   DB:                      13          Gaps:            0

   US-10-001-254-26 (1-59) x US-10-027-632-50724 (1-497)

   Qy    22 LeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLysLys 41
             |||   ||::|       |||  ||||||||
   Db   304 CTTCTAGATTTTGTGTCCCCTKACAAGGGTTGGAAGAGTCTGGCTCGAGCCCTGGCCTCT 245

   Qy    42 ProSerGly 44
             |||:||:|||
   Db   244 CCAGCTGGC 236

   RESULT 12
   US-10-027-632-50723/c
   Sequence 50723, Application US/10027632
   GENERAL INFORMATION:
   APPLICANT: Wang, David G.
   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
   TITLE OF INVENTION: Polymorphisms in the Human Genome
   FILE REFERENCE: 108827.129
   CURRENT APPLICATION NUMBER: US/10/027,632
   CURRENT FILING DATE: 2002-04-30
          PRIOR APPLICATION NUMBER: US 60/218,006
          PRIOR FILING DATE: 2000-07-12
          PRIOR APPLICATION NUMBER: US 60/198,676
          PRIOR FILING DATE: 2000-04-20
          PRIOR APPLICATION NUMBER: US 60/193,483
          PRIOR FILING DATE: 1999-11-23
          PRIOR APPLICATION NUMBER: US 60/167,363
          PRIOR FILING DATE: 2000-03-29
          PRIOR APPLICATION NUMBER: US 60/185,218
          PRIOR FILING DATE: 2000-02-24
          PRIOR APPLICATION NUMBER: US 60/156,358
          PRIOR FILING DATE: 1999-09-28
          PRIOR APPLICATION NUMBER: US 60/146,002
          PRIOR FILING DATE: 1999-08-09
          NUMBER OF SEQ ID NOS: 325720
          SOFTWARE: FastSEQ for Windows Version 4.0
   ;
   ;  SEQ ID NO 50723
   ;  LENGTH: 497
   ;  TYPE: DNA
   ;  ORGANISM: Human
   US-10-027-632-50723

   Alignment Scores:
   Pred. No.:               0.0477      Length:          497
   Score:                   73.00       Matches:         12
   Percent Similarity:      69.5%       Conservative:    4
   Best Local Similarity:   52.1%       Mismatches:      7
   Query Match:             22.60%      Indels:          0
   DB:                      14          Gaps:            0

   US-10-001-254-26 (1-59) x US-10-027-632-50723 (1-497)

   Qy    22 LeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleLysLys 41
             |||   ||::|       |||  ||||||||
   Db   304 CTTCTAGATTTTGTGTCCCCTKACAAGGGTTGGAAGAGTCTGGCTCGAGCCCTGGCCTCT 245

   Qy    42 ProSerGly 44
             |||:||:|||
   Db   244 CCAGCTGGC 236

   RESULT 13
   US-10-027-632-50724/c
   Sequence 50724, Application US/10027632
   GENERAL INFORMATION:
   APPLICANT: Wang, David G.
   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
   TITLE OF INVENTION: Polymorphisms in the Human Genome
   FILE REFERENCE: 108827.129
```

```
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE:
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   DB:
;   SEQ ID NO 50724
;   TYPE: DNA
;   ORGANISM: Human
;   LENGTH: 497
US-10-027-632-50724

Alignment Scores:
Pred. No.:
Score:                  0.0477
Percent Similarity:     73.00
Best Local Similarity:  69.57%
Query Match:            52.17%
                        22.60%
                        14

Length:        497
Matches:       12
Conservative:  4
Mismatches:    7
Indels:        0
Gaps:          0

US-10-001-254-26 (1-59) x US-10-027-632-50724 (1-497)

Qy  22 LeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleIleLysLys 41
       ||| ::|||||||||::|
Db 304 CTTCTAGATTTGTCTCCCCTKACAAGGGTTGGAAGAGTTCTGGCTGCAGCCCTGGCCTCT 245

Qy  42 ProSerGly 44
       |||::|||
Db 244 CCAGCTGGC 236

RESULT 14
US-10-027-632-72567/c
Sequence 72567, Application US/10027632
Publication No. US20030240075A9
;   GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   TITLE OF INVENTION: Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   DB:
;   SEQ ID NO 72567
;   TYPE: DNA
;   ORGANISM: Human
;   LENGTH: 499
US-10-027-632-72567
```

```
US-10-027-632-72567

Alignment Scores:
Pred. No.:
Score:                  0.048
Percent Similarity:     73.00
Best Local Similarity:  69.57%
Query Match:            52.17%
                        22.60%
                        13

Length:        499
Matches:       12
Conservative:  4
Mismatches:    7
Indels:        0
Gaps:          0

US-10-001-254-26 (1-59) x US-10-027-632-72567 (1-499)

Qy  22 LeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleIleLysLys 41
       ||| ::|||||||||::|
Db 306 CTTCTAGATTTGTCTCCCCTKACAAGGGTTGGAAGAGTTCTGGCTGCAGCCCTGGCCTCA 247

Qy  42 ProSerGly 44
       |||::|||
Db 246 CCAGCTGGC 238

RESULT 15
US-10-027-632-72568/c
Sequence 72568, Application US/10027632
Publication No. US20030240075A9
;   GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   TITLE OF INVENTION: Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   DB:
;   SEQ ID NO 72568
;   TYPE: DNA
;   ORGANISM: Human
;   LENGTH: 499
US-10-027-632-72568

Alignment Scores:
Pred. No.:
Score:                  0.048
Percent Similarity:     73.00
Best Local Similarity:  69.57%
Query Match:            52.17%
                        22.60%
                        13

Length:        499
Matches:       12
Conservative:  4
Mismatches:    7
Indels:        0
Gaps:          0

US-10-001-254-26 (1-59) x US-10-027-632-72568 (1-499)

Qy  22 LeuSerAspPheIleAspProGlnGluGlyTrpLysLysLeuAlaValAlaIleIleLysLys 41
       ||| ::|||||||||::|
Db 306 CTTCTAGATTTGTCTCCCCTKACAAGGGTTGGAAGAGTTCTGGCTGCAGCCCTGGCCTCA 247

Qy  42 ProSerGly 44
       |||::|||
Db 246 CCAGCTGGC 238

Search completed: January 18, 2004, 09:06:06
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:　　　　January 17, 2004, 23:09:11 ; Search time 1315 seconds
　　　　　　　　　　(without alignments)
　　　　　　　　　　5433.852 Million cell updates/sec

Title:　　　　　US-10-001-254-5
Perfect score:　294
Sequence:　　　1 acatatgtgcgctgcccaa..........tgctccagatgctgttccc 294
Scoring table:　IDENTITY_NUC
　　　　　　　　Gapop 10.0 , Gapext 1.0
Searched:　　　22781392 seqs, 1215223805 residues
Total number of hits satisfying chosen parameters:　45562784
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
　　　　　　　　Maximum Match 100%
　　　　　　　　Listing first 45 summaries

Database :　　　EST:*
　　1: em_estba:*
　　2: em_esthum:*
　　3: em_estin:*
　　4: em_estmu:*
　　5: em_estov:*
　　6: em_estpl:*
　　7: em_estro:*
　　8: em_htc:*
　　9: gb_est1:*
　10: gb_est2:*
　11: gb_htc:*
　12: gb_est3:*
　13: gb_est4:*
　14: gb_est5:*
　15: em_estom:*
　16: em_estfun:*
　17: em_estvr:*
　18: em_gss_hum:*
　19: em_gss_pln:*
　20: em_gss_vrt:*
　21: em_gss_fun:*
　22: em_gss_mam:*
　23: em_gss_mus:*
　24: em_gss_pro:*
　25: em_gss_rod:*
　26: em_gss_phg:*
　27: em_gss_vrl:*
　28: gb_gss1:*
　29: gb_gss2:*

SUMMARIES

Result No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	% Query Match Length	DB	ID	Description
1	294	100.0	859	10	BG164491	BG164491 602342026
2	292.4	99.5	811	10	BG616438	BG616438 602642772
3	267.2	90.9	719	10	BF696981	BF696981 602130160
4	262	89.1	402	10	BE482619	BE482619 168463 BA

ALIGNMENTS

RESULT 1
BG164491
BG164491
LOCUS　　　BG164491　　　859 bp　　mRNA　　linear　EST 06-FEB-2001
DEFINITION　602342026F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4452055 5',
　　　　　　　mRNA sequence.
ACCESSION　BG164491
VERSION　　BG164491.1　GI:12671194
KEYWORDS　EST.
SOURCE　　Homo sapiens (human)
ORGANISM　Homo sapiens
　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　1　(bases 1 to 859)
AUTHORS　NIH-MGC http://mgc.nci.nih.gov/.
TITLE　　National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL　Unpublished
COMMENT　Contact: Robert Strausberg, Ph.D.
　　　　　Email: cgapbs-r@mail.nih.gov
　　　　　Tissue Procurement: ATCC
　　　　　cDNA Library Preparation: Life Technologies, Inc.
　　　　　cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
　　　　　DNA Sequencing by: Incyte Genomics, Inc.
　　　　　Clone distribution: MGC clone distribution information can be
　　　　　found through the I.M.A.G.E. Consortium/LLNL at:
　　　　　http://image.llnl.gov
　　　　　Plate: LLAM10240　row: d　column: 08
　　　　　High quality sequence stop: 634.

5	220	74.8	541	10	BG691069	BG691069 340084 BA
5	205.4	69.9	313	12	BM151935	BM151935 TCBAP1811
6	204.4	69.5	503	14	CA538859	CA538859 C0272B01-
7	204.4	69.5	598	13	BQ552228	BQ552228
8	204.4	69.5	2810	11	AK029028	AK029028 Mus muscu
9	204.4	69.5	575	4	Bx522921	Bx522921 RZPD Mus
10	204.4	69.5	265	10	AW436511	AW436511 7674 MAR
11	204.4	69.5	507	10	BB866698	BB866698 BB866698
12	204.4	69.5	524	9	AW106160	AW106160 um23h11.y
13	204.4	69.5	453	10	BB860349	BB860349 BB860349
14	204.4	69.5	852	13	BU209111	BU209111 603950834
15	204.4	69.5	600	9	AJ453616	AJ453616 AJ453616
16	204.4	69.5	670	9	AJ699213	AJ699213 AJ699213
17	202.8	69.1	540	9	AJ447581	AJ447581 AJ447581
18	198.8	67.6	505	10	BB482323	BB482323 168064 BA
19	182.4	62.0	284	9	AL647125	AL647125 AL647125
20	181.4	61.7	629	9	AA114228	AA114228 zn7sg05.r
21	163.2	55.5	576	10	BF238344	BF238344 60190613
22	147.6	50.2	270	10	BB845841	BB845841 232716 BA
23	144.4	49.1	664	12	BJ035962	BJ035962 BJ035962
24	144.4	49.1	858	10	BF687921	BF687921 602069996
25	144.4	49.1	520	10	BE132064	BE132064 db41h09.y
26	147.6	50.2	453		AQ601806	AQ601806 HS_2108_B
27	139.8	47.6	632	14	CA365604	CA365604 640775 NC
28	137.6	46.8	795	13	BQ293310	BQ293310 EST01080
29	134	45.6	781	9	BF723681	BF723681 EST00673
35	125.8	42.8	555	9	AW423082	AW423082 fi68b07.y
36	123.4	42.0	743	14	CA474136	CA474136 AGENCOURT
37	113.6	38.6	380	12	BI883790	BI883790 fn07c09.y
38	107.4	36.5	490	10	AJ443799	AJ443799 AJ443799
39	106.2	36.1	595	28	AQ601806	AQ601806 HS_2108_B
40	79.6	27.1	335	13	BU430802	BU430802 UI-HF-BN0
41	69	23.5	873	28	BH150262	BH150262 ENTPF28TR
42	45.6	15.5	318	14	CB940191	CB940191 IpCGJx14_
43	44.2	15.0	491	9	AI802160	AI802160 tx25a02.x_
44	42.6	14.5				

FEATURES   Location/Qualifiers
     source   1..859
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /clone="IMAGE:4452055"
                  /db_xref="taxon:9606"
                  /tissue_type="hypernephroma, cell line"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NIH_MGC_89"
                  /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                  Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                  Average insert size 1.3 kb. Library enriched for
                  full-length clones and constructed by Life Technologies.
                  Note: this is a NIH_MGC Library."

BASE COUNT   266 a   168 c   190 g   235 t
ORIGIN

Query Match
Best Local Similarity   100.0%;   Score 294; DB 10;   Length 859;
Matches 294; Conservative 0%; Pred. No. 1.6e-71;
                         0; Mismatches 0; Indels 0; Gaps   0;

Qy   1 ACAATATGTCGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCTT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   59 ACAATATGTCGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 118

Qy   61 CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTAATTAAAAAAACCATCTGGTGATGATAGATAC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   119 CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTAATTAAAAAAACCATCTGGTGATGATAGATAC 178

Qy   121 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTAACTACTTCAAACTGAAAAAAGTCCCACTTCT 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   179 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTAACTACTTCAAACTGAAAAAAGTCCCACTTCT 238

Qy   181 GAATTACTGTTTGACTGGGGCACCACAAAATGCACAGTTGGTGATCTTGTGGGATCTTTTTG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   239 GAATTACTGTTTGACTGGGGCACCACAAAATGCACAGTTGGTGATCTTGTGGGATCTTTTTG 298

Qy   241 ATCCAAAAATGAAATTTTTCTCCTCGCGAGTCTTTTGCTCCCAGATAGCTGTTCCC 294
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   299 ATCCAAAAATGAAATTTTTCTCCTCGCGAGTCTTTTGCTCCCAGATAGCTGTTCCC 352

RESULT 2
LOCUS   BG616438   811 bp   mRNA   linear   EST 18-APR-2001
DEFINITION   602642772F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4773760 5',
               mRNA sequence.
ACCESSION   BG616438
VERSION   BG616438.1   GI:13667809
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 811)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: CLONETECH Laboratories, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLCM1645 row: p column: 17
               High quality sequence start: 3
               High quality sequence stop: 613.
FEATURES   Location/Qualifiers
     source   1..811
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /clone="IMAGE:4773760"
                  /db_xref="taxon:9606"
                  /tissue_type="embryonal carcinoma"
                  /clone_lib="NIH_MGC_61"
                  /lab_host="DH10B (T1 phage-resistant)"
                  /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                  SfiI (ggccgcctcggcc); Site 2: SfiI (ggccattatggcc);
                  Double-stranded cDNA was prepared from cell line RNA. 5'
                  and 3' adaptors were used in cloning as follows: 5' adaptor
                  sequence: 5'-CACGCGCCATTATGGCC-3' and 3' adaptor
                  sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACACTG-dT(30)BN-3';
                  (where B = A, C, or G and N = A, C, G, or T). Average
                  insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                  contained inserts by PCR. This library was enriched for
                  full-length clones and was constructed by Clontech
                  Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                  Library."

BASE COUNT   246 a   196 c   164 g   205 t
ORIGIN

Query Match
Best Local Similarity   99.5%;   Score 292.4; DB 10;   Length 811;
Matches 293; Conservative 99.7%; Pred. No. 4.4e-71;
                         0; Mismatches 1; Indels 0; Gaps   0;

Qy   1 ACAATATGTCGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCTT 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   77 ACAATATGTCGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 136

Qy   61 CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTAATTAAAAAAACCATCTGGTGATGATAGATAC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   137 CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTAATTAAAAAAACCATCTGGTGATGATAGATAC 196

Qy   121 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTAACTACTTCAAACTGAAAAAAGTCCCACTTCT 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   197 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTAACTACTTCAAACTGAAAAAAGTCCCACTTCT 256

Qy   181 GAATTACTGTTTGACTGGGGCACCACAAAATGCACAGTTGGTGATCTTGTGGGATCTTTTTG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   257 GAATTACTGTTTGACTGGGGCACCACAAAATGCACAGTTGGTGATCTTGTGGGATCTTTTTG 316

Qy   241 ATCCAAAAATGAAATTTTTCTCCTCGCGAGTCTTTTGCTCCCAGATAGCTGTTCCC 294
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   317 ATCCAAAAATGAAATTTTTCTCCTCGCGAGTCTTTTGCTCCCAGATAGCTGTTCCC 370

RESULT 3
LOCUS   BF696981   719 bp   mRNA   linear   EST 22-DEC-2000
DEFINITION   602130160F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4287014 5',
               mRNA sequence.
ACCESSION   BF696981
VERSION   BF696981.1   GI:11982389
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 719)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: CLONETECH Laboratories, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLCM1124 row: g column: 15
               High quality sequence stop: 632.
FEATURES   Location/Qualifiers
     source   1..811
                  /organism="Homo sapiens"

RESULT 4

```
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /tissue_type="primitive neuroectoderm"
                     /clone="IMAGE:4287014"
                     /db_xref="taxon:9606"
                     /clone_lib="NIH_MGC_56"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone="IMAGE:4287014"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgccctaggcc); Site_2: SfiI (ggccattaggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT       216 a    145 c    163 g    195 t
ORIGIN

Query Match  90.9%;  Score 267.2;  DB 10;  Length 719;
Best Local Similarity  98.3%;  Pred. No. 4.8e-64;
Matches 291;  Conservative 0;  Mismatches 3;  Indels 2;  Gaps 2;

Qy      1 ACATATGTGCGCCTGCCTCAANTGTTGGACTAATTAGGAAGCGTGTCAGATTTTATTGATCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 ACATATGTGCGCCTGCCTCAANTGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 151

Qy     61 CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATTCGGTGATGATAGATAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATTCGGTGATGATAGATAC 211

Qy    121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 271

Qy    181 GAATTACT-GTTTGACTGGGGCACACAAATTGCACAGTTGGTGGATCTTTGTTT 239
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    272 GAATTACTCGTTTGACTGGGGCACACAAATTGCACAGTTGGTGGATCTTTGTTT 331

Qy    240 GATCCAAAAATGAATT-TTTGCTTCCTGCGACGTCTTTTGCTCCCAGATGCTGTTCCC 294
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    332 GATCCAAAAATGAATTTTTTGCTTCCTGCGACGTCTTTTGCTCCCAGATGCTGTTCCC 387
```

LOCUS   BE482619   402 bp   mRNA   linear   EST 27-MAR-2003
DEFINITION 168463 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE482619
VERSION   BE482619.1 GI:9602152
KEYWORDS   EST.
SOURCE   Bos taurus (cow)
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 402)
  AUTHORS   Sonstegard,T., Capuco,A.V., White,J., Van Tassell,C.P., Connor,E.E.
            , Cho.,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D. and
            Quackenbush,J.
  TITLE   Analysis of bovine mammary gland EST and functional annotation of
            the Bos taurus gene index
  JOURNAL   Mamm. Genome 13 (7), 373-379 (2002)
  MEDLINE   22135956
   PUBMED   12140684
COMMENT   Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bdlg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416

RESULT 5

```
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="BARC 5BOV"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled mRNA isolated from mammary
                     tissues at eight physiological, developmental, and disease
                     states."
BASE COUNT       114 a     95 c     88 g    105 t
ORIGIN

Query Match  89.1%;  Score 262;  DB 10;  Length 402;
Best Local Similarity  93.2%;  Pred. No. 1.1e-62;
Matches 274;  Conservative 0;  Mismatches 20;  Indels 0;  Gaps 0;

Qy      1 ACATATGTGCGCCTGCCTCAANTGTTGGACTAATTAGGAAGCGTGTCAGATTTTATTGATCCT 60
          |||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     88 ACATACGTGCGGCTGCCTTCAGTGTTCTTGGACTAATTAGAAAAGCTGTTCCGATTTTATTGATCCT 147

Qy     61 CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATTCGGTGATGATAGATAC 120
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    148 CAAGAAGGATGGGAAGAAGTTAGCAGTCGCTATTAAAAAAACCATTCGGTGATGATAGATAC 207

Qy    121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    208 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 267

Qy    181 GAATTACTGTTTGACTGGGGCACACAAATTGCACAGTTGGTGGATCTTTGTTTG 240
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    268 GAGTTACTGTTTGACTGGGGCACCACCAAATTGCACAGTTGGTGAATCTTTTG 327

Qy    241 ATCCAAAAATGAATTTTTTTGCTTCCTGCGGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
          |||||||||| ||||||||||||||||||| |||||||||||||||||||||||||||
Db    328 GTCCAAAAATGAGTTTTTTTTGCCCCTGCAAGTCTTTTGCTACCAGATGCTGTACCC 381
```

LOCUS   BG691069   541 bp   mRNA   linear   EST 27-MAR-2003
DEFINITION 340084 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BG691069
VERSION   BG691069.1 GI:13932889
KEYWORDS   EST.
SOURCE   Bos taurus (cow)
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 541)
  AUTHORS   Sonstegard,T., Capuco,A.V., White,J., Van Tassell,C.P., Connor,E.E.
            , Cho.,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D. and
            Quackenbush,J.
  TITLE   Analysis of bovine mammary gland EST and functional annotation of
            the Bos taurus gene index
  JOURNAL   Mamm. Genome 13 (7), 373-379 (2002)
  MEDLINE   22135956
   PUBMED   12140684
COMMENT   Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center

Fax: 301 504 8414
Email: tads@lpsi.barc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 11 row: F column: 7
Seq primer: ATTTAGGTGACACTATAG.

RESULT 6

```
LOCUS       BM151935
DEFINITION  TCBAP1E11457 Pediatric pre-B cell acute lymphoblastic leukemia
            Baylor-HGSC project=TCBA Homo sapiens cDNA clone TCBAP1145, mRNA
            sequence.
ACCESSION   BM151935
VERSION     BM151935.1  GI:17175815
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313)
AUTHORS     Wei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman,F.R.Jr.,
            Gunatne,P.H., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F.
TITLE       Pediatric Leukemia cDNA Sequencing Project (2001)
JOURNAL     Unpublished
COMMENT     Contact: Dr. Judith F. Margolin
            Texas Children's Cancer Center and Human Genome Sequencing Center
            at Baylor College of Medicine
FEATURES             Location/Qualifiers
     source          1..313
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="TCBAP1145"
                     /sex="male"
                     /tissue_type="leukopheresis"
                     /cell_type="pre-B cell"
                     /dev_stage="pediatric 2 years"
                     /clone_lib="Pediatric pre-B cell acute lymphoblastic
                     leukemia Baylor-HGSC project=TCBA"
                     /note="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
                     First strand cDNA was primed with an anchored
                     XhoI-oligo(dT) primer [5'-GGAGGACTCGAGCGGCCGCCAGGAGGAG
                     3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand
                     was primed with a BamHI-dC primer
                     [5'-AGAGAGCTCGGATCCCGCGCGCGCGCAATAATAATAAT(C) 3'].
                     Double-stranded cDNA was then digested with BamHI and XhoI
                     and directionally cloned into the BamHI and SalI sites of
                     lambda pSB vector. Library went through one round of
                     normalization. Library was constructed by Wei Yu at RIKEN
                     of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,
                     Itoh M, Nagaoka S, Sasakin, Okazaki Y, Muramatsu M,
                     Schneider C, Hayashizaki Y, High efficiency selection of
                     full-length cDNA by improved biotinylated cap trapper.,
                     DNA Res 4: 1, 61-6, Feb 28, 1997)"
BASE COUNT       82 a     62 c     81 g     87 t    1 others
ORIGIN
```

Query Match   69.9%; Score 205.4; DB 12; Length 313;
Best Local Similarity   92.7%; Pred. No. 7e-47;
Matches 215; Conservative   0; Mismatches 17; Indels   0; Gaps 0;

```
Qy     1  ACATATGTGGGGCTGCCTCAATGTTGGACTAATTAGGAAGCTTCAGATTTTATTGATCCT   60
          ||||| ||||| || ||||||||| ||||||||||||||||| ||||||| |||||| ||
Db    77  ACATATGTGGGCTGCCTCAATGTTGGACTAATTAGGAAGCTGCAGATTTTATTGATCCT  136

Qy    61  CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTCGGTGATGATAGATAC  120
          ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   137  CAAGAAGGATCGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTCGGTGATGATAGATAC  196

Qy   121  AATCAGTTTCACATAAGGAGGAGATTTGAAAGCATTACTTCAAAACTGGGAAAAAGTCCACTTCT  180
          |||||||||||||||||||||||||||||| |||| ||||||||||||||||||| |||||||
Db   197  AATCAGTTTCACATAAGGAGGAGATTTGAAGCATTACTTCTNACTGGGAAAAAGTCCCACTTCT  256

Qy   181  GAATTACTGTTTGACTGGGGCGCCACCAAAAATTGCACAGTTGGTGATCTTGTGTGG  232
          |||||||||||||||||||| ||||||||| |||||||||||| ||| ||||||
Db   257  GAATTACTGTTTGACTGGGGCGCGGCCGCGCTCTTGGGCAGTTGTTTGTGTGG  308
```

RESULT 7

```
LOCUS       CA538859
DEFINITION  C0272B01-5N NIA Mouse 7.5-dpc Whole Embryo cDNA Library (Long) Mus
            musculus cDNA clone NIA:C0272B01 IMAGE:30017484 5', mRNA sequence.
ACCESSION   CA538859
VERSION     CA538859.1  GI:25080828
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 503)
AUTHORS     Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Tanaka,T., Lim,M.K.,
            Luo,A. and Ko,M.S.H.
TITLE       Systematic Analyses of NIA Mouse 7.5-dpc Whole Embryo cDNA Library
```

```
LOCUS       BM151935
DEFINITION  TCBAP1E11457 Pediatric pre-B cell acute lymphoblastic leukemia
            Baylor-HGSC project=TCBA Homo sapiens cDNA clone TCBAP1145, mRNA
            sequence.                      313 bp    mRNA    linear    EST 30-NOV-2001
```

FEATURES   Location/Qualifiers
     source   1..541
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="BARC 5BOv"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled mRNA isolated from mammary
                     tissues at eight physiological, developmental, and disease
                     states."

```
BASE COUNT      169 a     96 c    119 g    157 t
ORIGIN
```

Query Match   74.8%; Score 220; DB 10; Length 541;
Best Local Similarity   93.9%; Pred. No. 7.2e-51;
Matches 229; Conservative   0; Mismatches 15; Indels   0; Gaps 0;

```
Qy    51  TATTGATCCTCAGGAAGGATGGGAAGAAGTTAGCGTGTAGCTATTAAAAAACCATCTCGGTGA  110
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  TATTGATCCTCAGGAAGGATGGGAAGAAGTTAGCAGTGTCGCTATTAAAAAACCATCTGGGTGA   60

Qy   111  TGATAGATACAATCAGTTTCACATAAGGAGGAGATTTGAAGCATTACTTCAAACTGGGAAAAG  170
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    61  TGATAGATACAATCAGTTTCACATAAGGAGGAGATTTGAAGCATTACTGCAAATTGGAAAAAG  120

Qy   171  TCCCACTTCTGAATTACTGTTTGACTGGGGGCGGCCACCAAAAATTGCACAGTTGGTGATCTTGT  230
          |||||||||||||||||||||||||||| || ||||||||||||||||||||||||| ||||
Db   121  CCCCAGCTGTGAGTTACTGTTTGACTGGGGCCACCACAAAAATTGCACAGTTGGTGATCTTGT  180

Qy   231  GGATCTTTGATCCAAAAATGAAATTTTTTGCCTCCGCGAGTCTCTTTTTGCTCCCAGATGCTGT  290
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   181  GGATTATTTGGTCTCAAAATGAGTTTTTTTCCCCTGCGAGTCTCTTTTTGCTACCAGATGGCTGT  240

Qy   291  TCCC                                                          294
          ||||
Db   241  ACCC                                                          244
```

JOURNAL (Long) Unpublished
COMMENT Other_ESTs: C0272B01-3
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: C0272 row: B column: 01
Seq_primer: M13 Reverse
High_quality_sequence_stop: 503
POLYA=No.

FEATURES   Location/Qualifiers
source   1..503
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/clone="NIA:C0272B01 IMAGE:30017484"
/db_xref="taxon:10090"
/db_xref="niaEST:C0272B01-5N"
/tissue_type="whole embryo including extraembryonic tissues at 7.5-days postcoitum"
/dev_stage="7.5-days postcoitum"
/clone_lib="NIA Mouse 7.5-dpc whole Embryo cDNA library
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11541199]). Total RNAs were extracted from a pool of four embryos at 7.5-days postcoitum. Double-stranded cDNAs were synthesized with an Oligo(dT) primer (Invitrogen). 5'-pGACTAGTTCTAGATCGGACGGGCCCGCCCCCTTTTTTTTTTTTTTT-3'] from 7 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-Linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT   119 a   146 c   138 g   100 t
ORIGIN

Query Match 69.5%; Score 204.4; DB 14; Length 503;
Best Local Similarity 81.0%; Pred. No. 1.6e-46;
Matches 238; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
Qy    1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGAGCTGTCAGATTTTATTGATCCT   60
        ||||||||| |||||||||||||||| |||||||||||||||||||||||||||||||||||
Db  192 ACATACATACGCAACCTTAATGTGGGGATCCTTAGGAAGCTGTCGGATTTTATTGATCCT  251

Qy   61 CAAGAAGGATGGAAGAAAGTTAGCTGTAGCTATTAAAAACCATCTGGTGATGATAGATAC  120
        ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  252 CAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAAAAAGCCGTCCCGCGACGACAGATAC  311

Qy  121 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  312 AATCAGTTTCCATATAAGGAGAGATTCAGAGCCTACTTCAGACCGGGAAGAGAGCCCACCTGT  371

Qy  181 GAATTACTGTTTGACTGGGGGCACACAAAATTGCACAGTTGGTGATCTTGTTGTGGATCTTTTG  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  372 GAACTGCTGTTTGACTGGGGGCACCACGGAACTGCACAGTTGGCGGACCTTGTGGGATCTACTG  431

Qy  241 ATCCAAAATGAATTTTTGCTCCTCGGAGGTCTTTGCTCCGAGATGCTGTTCCC  294
        | |||||||||
Db  485 GTCCAGATTGAGCTGTTGTTGCCCCGCGCACTGCCACTCTCCCTGCTGCTGCCGGCGATGCGCGTTCCC  485
```

Db 432 GTCCAGATTGAGCTGTTGTTGCCCCGCGCACTCTCCCTGCTGCTGCCGGCGATGCGCGTTCCC 485

RESULT 8
LOCUS   BQ552228   598 bp   mRNA   linear   EST 20-JUN-2002
DEFINITION H4014C09-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone H4014C09 5', mRNA sequence.
ACCESSION BQ552228
VERSION BQ552228.1 GI:21453114
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 598)
AUTHORS VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G., Martin ,P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T., Kargul,G.J., Luo,A.G., Kelso,J., Hide,W. and Ko,M.S.H.
TITLE Assembly, verification, and initial annotation of NIA 7.4K mouse cDNA clone set
JOURNAL Genome Res. 12 (12), 1999-2003 (2002)
MEDLINE 22354164
PUBMED 12466305
COMMENT Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please visit http://lgsun.grc.nia.nih.gov/cDNA/NIA_7_4k.html for details.
Plate: H4014 row: C column: 09
Seq_primer: -21M13 Reverse
High_quality_sequence_stop: 598
POLYA=No.

FEATURES   Location/Qualifiers
source   1..598
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="niaEST:H4014C09-5"
/db_xref="taxon:10090"
/clone="H4014C09"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/clone_lib="NIA Mouse 7.4K cDNA Clone Set"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 7,407 clones from more than 20 cDNA libraries."
POLYA=No.

BASE COUNT   150 a   174 c   163 g   111 t
ORIGIN

Query Match 69.5%; Score 204.4; DB 13; Length 598;
Best Local Similarity 81.0%; Pred. No. 1.8e-46;
Matches 238; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
Qy    1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGAGCTGTCAGATTTTATTGATCCT   60
        ||||||||| |||||||||||||||| |||||||||||||||||||||||||||||||||||
Db  176 ACATACATACGCAACCTTAATGTGGGGATCCTTAGGAAGCTGTCGGATTTTATTGATCCT  235

Qy   61 CAAGAAGGATGGAAGAAAGTTAGCTGTAGCTATTAAAAACCATCTGGTGATGATAGATAC  120
        ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  236 CAAGAAGGGTGGAAGAAATTAGCAGTAGCTATCAAAAAAGCCGTCCCGCGACGACAGATAC  295

Qy  121 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  296 AATCAGTTTCCATATAAGGAGAGATTCAGAGCCTACTTCAGACCGGGAAGAGAGCCCACCTGT  355

Qy  181 GAATTACTGTTTGACTGGGGGCACACAAAATTGCACAGTTGGTGATCTTGTTGTGGATCTTTTG  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  356 GAACTGCTGTTTGACTGGGGGCACCACGGAACTGCACAGTTGGCGGACCTTGTGGGATCTACTG  415
```

```
Qy      241 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCCCAGATGCTGTTCCC 294
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      416 GTCCAGATTGAGCTGTTTGCCCCGCCCGGCCACTCTCCTGCGCTGCCGGTTCC 469
```

RESULT 9

BB660378

LOCUS   BB660378   610 bp   mRNA   linear   EST 26-OCT-2001
DEFINITION   BB660378 RIKEN full-length enriched, 13 days embryo lung Mus
                musculus cDNA clone D43004L21 5', mRNA sequence.
ACCESSION   BB660378
VERSION   BB660378.1 GI:16494199
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 610)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
                , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
                ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
                Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
                ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
                Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
                Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT   Contact: Yoshihide Hayashizaki
                Laboratory for Genome Exploration Research Group, RIKEN Genomic
                Sciences Center(GSC), Yokohama Institute
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81-45-503-9222
                Fax: 81-45-503-9216
                Email: genome-res@gsc.riken.go.jp,
                URL:http://genome.gsc.riken.go.jp/
                Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
                ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                Normalization and subtraction of cap-trapper-selected cDNAs to
                prepare full-length cDNA libraries for rapid discovery of new
                genes. Genome Res. 10 (10), 1617-1630 (2000)
                Watahiki,M., Yoneda,Y., Izawa,M., Ohara,E.,
                wagi,K., Fujiwake,S., Inoue,Y., Togawa,Y.,
                ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
                Hayashizaki,Y.
                RIKEN integrated sequence analysis (RISA) system-384-format
                sequencing pipeline with 384 multicapillary sequencer. Genome Res.
                10 (11), 1757-1771 (2000)
                Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
                ,Y. and Hayashizaki,Y.
                Computer-based methods for the mouse full-length cDNA
                encyclopedia: real-time sequence clustering for construction of a
                nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
                Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
                ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
                Hayashizaki,Y.
                Computational Analysis of Full-Length Mouse cDNAs Compared with
                Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
                Please visit our web site (http://genome.gsc.riken.go.jp) for
                further details.
FEATURES   Location/Qualifiers
     source   1..610
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="D43004L21"
                     /tissue_type="lung"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     lung"

BASE COUNT.
ORIGIN
        149 a   179 c   164 g   118 t

```
Qy      1 ACATATGTCGCGGCTGCCTCAATGTTGGACTAATTAGGAAGCTGCTGTCAGATTTTATTGATCCT 60
            |||||||||||||||||||||  ||||| || |||| ||| ||||||| ||| || |||||||||
Db      212 ACATACATACGCCACCTTAATGTGGGAATCCTTAGGAGCGTGTCGGATTTTATTGATCCT 271

Qy      61 CAAGAAGGATGGAAGAAGTTAGCTGTTAGCTATTAATTAAAAAACCATCTGGTGATTGATAGATAC 120
            |||||||||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||
Db      272 CAAGAAGGGTGGGAGAAAATTAGCAGTAGCTATTCAAAAAGCGGCGCAGCCGACGACAGATAC 331

Qy      121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTTCAAACTGGAAAAAGTCCCACACTTCT 180
            ||||||| || |||||||||| ||||| |||||| ||||||||| ||||||| |||||||| ||
Db      332 AATCAGTTTCCATATAAGGAGAATTCGAAAGCCTTACTTACTTCAGACCGGGAAGAGCCCAACCTGT 391

Qy      181 GAATTACTTGTTTGACTTGGGGCACCACAAATTGCACAGTTTGGTGATCTTTTTG 240
            ||||| |||||||||||||||||||||||| || | |||||| ||| ||||||
Db      392 GAACTGCTGTTTGACTGGGGCCACCACGACTTGCGGATCTACTG 451

Qy      241 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCCCAGATGCTGTTCCC 294
            |||||||||||||||||||||||||||||||||||||||||||| |||||
Db      452 GTCCAGATTGAGCTGTTTGCCCCGCCCGGCCACTCTCCTGCGCTGCCGGTTCC 505
```

RESULT 10

BB613447

LOCUS   BB613447   637 bp   mRNA   linear   EST 26-OCT-2001
DEFINITION   BB613447 RIKEN full-length enriched, 10 day neonate skin Mus
                musculus cDNA clone 4732482P03 5', mRNA sequence.
ACCESSION   BB613447
VERSION   BB613447.1 GI:16454055
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 637)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
                , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
                ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
                Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
                ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
                Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
                Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT   Contact: Yoshihide Hayashizaki
                Laboratory for Genome Exploration Research Group, RIKEN Genomic
                Sciences Center(GSC), Yokohama Institute
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81-45-503-9222
                Fax: 81-45-503-9216
                Email: genome-res@gsc.riken.go.jp,

Query Match   69.5%; Score 204.4; DB 10; Length 610;
Best Local Similarity 81.0%; Pred. No. 1.8e-46;
Matches 238; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

Watahiki,A., Fujiwake,K., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES   Location/Qualifiers

source   1..637
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="4732482P03"
    /sex="mixed"
    /tissue_type="skin"
    /dev_stage="10 days neonate"
    /clone_lib="RIKEN full-length enriched, 10 day neonate skin"
    /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGAGAGAGAGACTCTTTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of 3'). cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

BASE COUNT   162 a   185 c   173 g   117 t
ORIGIN

Query Match 69.5%; Score 204.4; DB 10; Length 637;
Best Local Similarity 81.0%; Pred. No. 1.8e-05;
Matches 238; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```
Qy      1 ACAPATAGTGCGCTGCCTCCAATGTGTTGGACATTAAATTAGGAGACGTGTGACATTTTATTGATCCT 60
          |||||||||| |||| ||| ||||||||||||||| |||||||||||||| || ||||||||||||||
Db    193 ACATACAATACGCAACCTTAAATGTGGGGAATCCTTAGGAAGCTGTCGGATTTTATTGATCCT 252

Qy     61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGGTCGATGATAGAATAC 120
          ||||| || |||| ||||||||||||||||| |||||| | || ||||||| |||| | |||||
Db    253 CAAGAAGGGTGGAAGAAATTAGCAGTAGCTATTCAAAAAGCCGTCTCGGGCGACGACAGATAC 312

Qy    121 AATCAGTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAGTCCCACTTCT 180
          ||||||||| |||| ||||||| || |||| ||||| |||||| | ||| |||||| |
Db        
```

RESULT 11
BY721552

LOCUS   BY721552   638 bp   mRNA   linear   EST 17-DEC-2002
DEFINITION   BY721552 RIKEN full-length enriched, adult male diencephalon Mus musculus cDNA clone 9330209D03 5', mRNA sequence.
ACCESSION   BY721552
VERSION   BY721552.1 GI:27134669
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 638)
AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H. , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S. , Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A. , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J. , Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K. , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE   Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
MEDLINE   22354683
PUBMED   12466851
COMMENT   Contact: Yoshihide Hayashizaki
    Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
    The Institute of Physical and Chemical Research (RIKEN)
    1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
    Tel: 81-45-503-9222
    Fax: 81-45-503-9216
    Email: genome-res@gsc.riken.go.jp;
    URL:http://genome.gsc.riken.go.jp/
    Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K. , Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,N., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
    Computational Analysis of Full-Length Mouse cDNAs Compared with

```
Db    313 AATCAGTTCCAPATATAGGAGAGATTCGPAAGCCTTCACGACACGGGAAGAGCCCACCTGT 372
          ||||||||| | ||||| ||||||  || ||||| ||  | |||||| || |||| || |
Qy    181 GAATTACGTTTGACTGGGGCACCACAAAATTGCACAGTTGGTGTGATCTTGGGATCTTTTG 240
Db    373 GAACTGCGTTTGACTGGGGGATCCGACCACCACGAACTGCACAGTTGGCGGACCTTGTGGGATTCACTG 432
          |||| | |||||||||||||     | ||| ||||  |||||||||| |||  || |  |||| |||
Qy    241 ATCCAAAATGAATTTTTTGCCCTGCGAGTCTTTTGTCCCAGATGCTGTTCCC 294
Db    433 GTCCAGATTGAGCTGTTTTGCCCCGCCCCACTCTTCCTCGTCCCGGATGCGGTTCC 486
```

Human Genome Sequences Mamm. Genome, 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

```
FEATURES             Location/Qualifiers
source          1..638
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="9330209D03"
                /sex="male"
                /tissue_type="diencephalon"
                /dev_stage="adult"
                /lab_host="DH10B"
                /clone_lib="RIKEN full-length enriched, adult male
                diencephalon"
                /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                prepared and sequenced in Mouse Genome Encyclopedia
                Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in
                RIKEN. Division of Experimental Animal Research in Riken
                contributed to prepare mouse tissues. 1st strand cDNA was
                primed with a primer [5'
                GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                prepared by using trehalose thermo-activated reverse
                transcriptase and subsequently enriched for full-length by
                cap-trapper. cDNA went through one round of normalization
                to Rot = 10.0 and subtraction to Rot = 185.0. Second
                strand cDNA was prepared with the primer adapter [5'
                GAGGAGAGAGGATCCTCGAGTTAAATTAAAATTAAATCCCCCCCCCCCC
                3']. cDNA was cloned into the XhoI and BamHI sites.
                Vector: a modified pBluescript KS(+) after bulk excision
                from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                BamHI"
BASE COUNT      160 a    189 c    172 g    117 t
ORIGIN
```

Query Match   69.5%;   Score 204.4;   DB 14;   Length 638;
Best Local Similarity   81.0%;   Pred. No. 1.8e-46;
Matches 238;   Conservative 0;   Mismatches 56;   Indels 0;   Gaps 0;

```
Qy       1  ACAATATGTCGGCTGCCTCAATGTTGGACTAATTAGGAGCGTCAGATTTTTATTGATCCT   60
            ||||  || |||  | ||||||||||||| |||||||||||||| |||||||||||||||||
Db     198  ACATACATACGCAACCTTAATGTGGGGACCCTTAGGAGACCTGTCGGATTTTTATTGATCCT  257

Qy      61  CAAGAAGGATGGAGAAGAAGTTAGCTGTTAGCTATTAAAAAACCATCTGGTGATGAGATAC  120
            ||||||||   ||||||||| |||| |||| | ||||| |||| ||||| |||  |||||||
Db     258  CAAGAAGGGCTGGAAGAAAATTAGCAGTAGCTATCAAAAAGCCGTCCGGCGACGACAGATAC  317

Qy     121  AATCAGATTTCACATAAGCAGATTGTTGAAGCATTACTTCAAACTGAAAAAGTCCCACTTCT  180
            |||||| ||  ||||||| ||||| | |||||| ||||||||  ||| ||      |||  ||
Db     318  AATCAGTTCCATATAAGGAGATTCGAAGCCTTACTTCACAGTTCAG              377

Qy     181  GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGGTGGATCTTTTG         240
            |||  |||||||||||||||| |||| || ||||||||||||||   |||| ||
Db     378  GAACTGCGCTGTTTGACTGGGGCGCCACCACGAACTGCACAGTTGGGGACCTTACTG  437

Qy     241  ATCCAAAAATGAATTTTTTGCTCCTCGCGAGTGCTTTTGCTCCCAGATGCTGTTCCC      294
            |||  ||| ||| | ||||||| ||| |||||||||   |  ||||||| |||||||
Db     438  GTCCAGATTGAGCTGTTTCCCCCGCGCCCACTCTCTGCTCGCCGGATGCCGTTCCC  491
```

Db   438 GTCCAGATTGAGCTGTTTCCCCCGCGCCCACTCTCTGCTCGCCGGATGCCGTTCCC 491

```
Db
RESULT 12
LOCUS       BY726858                 663 bp    mRNA    linear    EST 17-DEC-2002
DEFINITION  BY726858 RIKEN full-length enriched, adult male corpora
            quadrigemina Mus musculus cDNA clone B230034I03 5', mRNA sequence.
ACCESSION   BY726858
VERSION     BY726858.1  GI:27139985
KEYWORDS    EST.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.,
            Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.,
            Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
   PUBMED   12466851
  MEDLINE   22354683
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sato,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome, 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
```

encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES   Location/Qualifiers
    source   1..663
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="B230034I03"
                     /sex="male"
                     /tissue_type="corpora quadrigemina"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male corpora
                     quadrigemina"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5',
                     GAGAGAGAGAGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter [5'
                     sequence [5' GAGGAGGAGATCTCGAGTTAATTAAATTAAATTCCCCCCCCCCCC
                     3'].cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT   165 a   190 c   181 g   124 t   3 others
ORIGIN

RESULT 13

Query Match
Best Local Similarity   69.5%;   Score 204.4; DB 14; Length 663;
Matches 238; Conservative   0; Mismatches 56; Indels 0; Gaps 0;

Qy   1 ACATATGTGCGCTGCCTCAATGTTCGGACTAATTAGGAGAGTGTCCAGATTTTATTGATCCT 60
          |||||| |||||| ||||||||| || |||| |||| |||| || |||||| |||| ||||
Db   185 ACATACATACGGCAACCTTAATGTGGGGATCCTTAGGAAGCTGTCGGATTTTATTGATCCT 244

Qy   61 CAAGAAGGATGGAAGAAGTTAGCTCTAGCTATTAAAAAAACCATCTGGGTGATGATGATAC 120
          |||||| || || |||| ||||||||| |||| ||||||||| || |||| |||||| ||
Db   245 CAAGAAGGGTGGGAAGAAAATTAGCAGTAGCTATCAAAAAGCCGTCCGCGACGACAGATAC 304

Qy   121 AATCAGTTTCACATAAGGAGATTTGAAGCAATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
          ||||| ||| ||||||||||||| || |||| ||||| |||| |||| ||| |||| |||
Db   305 AATCAGTTCCATATAAGGAGATTCGAAGCCTTACTTACTCAGACCGGAAGAGCCCCACCTGT 364

Qy   181 GAATTACTGTTTTGACTGGGGCCACCACACAAAATTGCACAGTTGGGTGATCTTGGGATCTTTTTG 240
          ||||| |||||||| || |||||| |||| |||| |||| |||| || || ||||||||
Db   365 GAACTGCTGTTTGACTGGGGCCACCACGAACTGCACAGTTGGCGACCTTGTCGGATCTACTG 424

Qy   241 ATCCAAAAATGAAATTTTTGCTCCTGCGAGATGCGTGTTCCC 294
          ||| |||| |||||||||| |||||||| |||||||||||||
Db   425 GTCCAGATTGAGCTGTTGCCCCCCGCCCACTCCTCCGGAGCGCGTTCCC 478

FEATURES   Location/Qualifiers
    source   1..676
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="473246O109"
                     /sex="mixed"
                     /tissue_type="skin"
                     /dev_stage="10 days neonate"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 10 day neonate
                     skin"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5',
                     GAGAGAGAGAGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization

```
RESULT 14
LOCUS       AK020397                1161 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male diencephalon cDNA, RIKEN full-length
            enriched library, clone:9330205003 product:interleukin-1
            receptor-associated kinase 4 [Mus musculus], full insert sequence.
ACCESSION   AK020397
VERSION     AK020397.2  GI:15919911
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99272953
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   2030913
```

Qy   1 ACATATGTGCGCCTGCCTCAATGTTGGACTTAAATTAGGAAGCTGTCAGATTTTATTGATCTT 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140 ACATATACAGCAACCTTAATGTGGGGGATCCTTAGGAAGCCTGTCGGATTTTATTGATCCTT 199

Qy   61 CAAGAAGGATGGAGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 200 CAAGAAGGGTGGAGAAGAAAATTAGCAGTAGCTATCAAAAACCGTTCGGCGACGACGAGATAC 259

Qy 121 AATCAGTTTCACATAATAAGAGAGATTTGAAGCATTACTTCAAAACTGGAAAAAGTCCCACTTCTT 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 260 AATCAGTTCCATATATAAGGAGATTCGAAGCCTTACTTCAGAGCCCGGAAGAGCCCACCTCGT 319

Qy 181 GAATTACTNGTTTGACTGGGGCGCACACAAAAATTGCACAGAGTTGGTGTGGATCTTTGGAGTCTTTTG 240
        |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 320 GAACTGCGTGTTGACTGGGGGCCACCACAGAACTGCACAGTTGTGGGACCTGTACTG 379

Qy 241 ATCCAAAATGAAATTTTTGCTCCTCGGAGTCTTTTGCTCCCAGATGCTGTGTTCCC 294
        ||||| ||| ||| |||||||||||||||||||||||||||||
Db 380 GTCCAGATTGAGCTGTTTGCCCCCCGGCCACTCTCCTCCTCGGATGCCGTTCCC 433

```
                         /mol_type="mRNA"
                         /strain="C57BL/6J"
                         /db_xref="FANTOM_DB:9330209D03"
                         /db_xref="MGI:1911393"
                         /db_xref="taxon:10090"
                         /clone="9330209D03"
                         /sex="male"
                         /tissue_type="diencephalon"
                         /clone_lib="RIKEN full-length enriched mouse cDNA library"
                         /dev_stage="adult"
     CDS             183..776
                         /codon_start=1
                         /product="unnamed protein product; interleukin-1
                         receptor-associated kinase 4 [Mus musculus]"
                         /note="unnamed protein product; interleukin-1
                         receptor-associated kinase 4 [Mus musculus]"
                         /protein_id="BAB32090.2"
                         /db_xref="GI:15919912"
                         /db_xref="MGI:1924812"
                         /translation="MNKPLTPSTYIRNLNVGILRKLSDFIDPQEGWKKLAVAIKKPSG
                         DDRXNQFHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLLVQIELFAPATLLP
                         DAVPQTVKSLPPKEAGNSGTNTRALSGKGQDIRNAYAKTRTQLRATGLLSPDNRSVES
                         SDFRFHSFSFHELKSITNNFDEQPASAGNRMGEGGF"
BASE COUNT      317 a     296 c     308 g     240 t
ORIGIN

Query Match     69.5%;  Score 204.4; DB 11; Length 1161;
Best Local Similarity   81.0%;  Pred. No. 2.4e-46;
Matches 238;  Conservative 0;  Mismatches 56;  Indels 0;  Gaps 0;

Qy    1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATTCCT 60
         |||||| ||| |||||||| |||||||||||||||||||||||||| ||| || ||||||||
Db  207 ACATACATACGCAACCTTAATGTGGGGATCCTTAGGAAGCTGTCGGATTTTATTGATTCCT 266

Qy   61 CAAGAGGATGGAAGAAGAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
         ||||||| ||| |||||| |||||| ||||||| |||| |||| |||||| |||| ||||
Db  267 CAAGAGGGTGGGAAGAAAATTAGCAGTAGCTATCAAAAAGCCGTCCGGCGACGACAGATAC 326

Qy  121 AATCAGTTTCACATAAGGAGGATTTGAAGCATTACTTCAAACTGTGGAAAAAGTCCCACTTCT 180
         |||| ||| || | |||| ||||| ||||  |||| || || ||||| ||| |||||||||
Db  327 AATCAGTTCCATATAAGGAGGATTCCGAAGCCTTACTTCAGACCGGGAAGAGAGACCCCACCTGT 386

Qy  181 GAATTACTGTTTGACTGGGGGCCACCACAAAATTGCACAGTTGGTTGGTGATCTTGTTGGGATCTTTTG 240
         ||||| | | ||||||||| |||||||||| ||||||||| ||| ||||||||| |||| ||||
Db  387 GAACTGCTGTTTTGACTGGGGCACCACCAGTTGGCGACCTTGTTGGGGATCTACTG 446

Qy  241 ATCCAAAAATGAAATTTTTGCTCTCCGAGTGTCTTTTGCTCCCAGATGCTGTTCCC 294
         ||||| ||||||||| || ||||||||||  ||||  ||||| |||| |||||||||
Db  447 GTCCAGATTGAGCTGTTTTTGCCCCCCGCCACTCTCTGCCGGATGCCGGTTCCC 500
```

RESULT 15

```
LOCUS       AK028837                2481 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
            library, clone:4732460I09 product:interleukin-1 receptor-associated
            kinase 4 [Mus musculus], full insert sequence.
ACCESSION   AK028837
VERSION     AK028837.1  GI:26324783
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,M.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C.A., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2481)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagami,M., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
```

```
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
```

```
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/.
URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2481
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:4732460I09"
                     /db_xref="taxon:10090"
                     /clone="4732460I09"
                     /tissue_type="skin"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     CDS             116..1495
                     /note="unnamed protein product; interleukin-1
                     receptor-associated kinase 4 [Mus musculus]
                     putative"
                     /codon_start=1
                     /protein_id="BAC26146.1"
                     /db_xref="GI:26324784"
                     /translation="MNKPLTPSTYIRNLNVGILRKLSDFIDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLLVDLLVQIELFAPATLLLP
                     DAVPQTVKSLPPREAATVAQTHGPQEEKDRTSVMPMKLEHSCEPPDSSSPDNRSVES
                     SDTRFHSFSFHELKSITNNFDEQPASAGQNRMGEGGFGVVYKGCVNNTIVAVKKLGAM
                     VEISTEELKQQFDQEIKVWATCQHENLVELLGFSSDSDNLCIVYAYMPNGSLLDRLSC
                     LDGTPPLSWHTRCKVAQGTANGIRFLHENHHIHRDIKSANILLDKDFTAKISDFGLAR
                     ASARLAQTVWTWTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLELITGLAAVDENRE
                     PQLLLDIKEEIEDEEKTIEDYTDEKMSDADPASVEAMYSAASQCLHEKKNRRPDIAKV
                     QQLLQEMSA"
BASE COUNT      650 a    592 c    634 g    605 t
ORIGIN
```

```
Query Match        69.5%;   Score 204.4;  DB 11;  Length 2481;
Best Local Similarity   81.0%;   Pred. No. 3.4e-46;
Matches 238;   Conservative    0;   Mismatches  56;   Indels   0;   Gaps    0;

Qy       1 ACATATNGTGGGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTTATTGATCCT    60
           ||||| |||| || ||||||||||||||||||||||||||||||||||||||||||||||||
Db     140 ACATACATACGCAACCTTAATGTGGGGGATCCTTAGGAAGCTGTCCGATTTTATTGATCCT   199

Qy      61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC   120
           ||||| || ||| ||||||||||||| |||||||||||| ||||||||| || ||||||||
Db     200 CAAGAAGGGTGGAAGAAAATTAGCAGTAGCTATCAAAAAGCCGTCCGGCGACGACAGATAC   259

Qy     121 AATCAGTTTCACATAAGGAGGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT   180
           |||||||| || ||||||||| ||| ||||||||||||||| |||||| |||||||||| |
Db     260 AATCAGTTCCATATAAGGAGGATTCGAAGCCCTTACTTCAGACCGGGAAGAGAGCCCCACCTGT   319

Qy     181 GAATTACTGTTTGACTGGGGCCACCACCAAATTGCACAGTTGGTTGGTTGTGGATCTTTTG   240
           |||||||||||||||||| ||| || ||||| |||||||||| ||||||| ||||| |  |
Db     320 GAACTGCTGTTTTGACTGGGGCCACCAGTTGGCCGACCTTTGTGGATCTTACTG   379

Qy     241 ATCCAAAATGAAATTTTTGCTCCTGCCAGTGTCTTTTGCTCCCAGATGCTGTTTCCC   294
           ||||| | ||||| | ||||| |  ||  |  || ||||||||||||||| |||||||
Db     380 GTCCAGATTGAGCTGTTTTGCCCCGCCACTCTCCTGCTGCCGGGGATGCCGGTTCCC   433

Search completed: January 18, 2004, 00:39:20
Job time : 1319 secs
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 17, 2004, 20:02:06 ; Search time 191 Seconds
(without alignments)
4155.156 Million cell updates/sec

Title: US-10-001-254-5
Perfect score: 294
Sequence: 1 acatatgtgcgctgcccaa..........tgctccagatgctgtccc 294

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 252756 seqs, 134919017 residues

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Total number of hits satisfying chosen parameters: 5105512

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
   /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

No.					ID	Description
9	150.4	51.2	405	23	AAS76803	DNA encoding novel
10	137.6	46.8	211	24	AAD40084	Human IRAK4 short
11	137.6	46.8	415	23	AAS76802	DNA encoding novel
12	137.6	46.8	2213	22	AAH13798	Human cDNA sequenc
13	104.8	35.6	408	24		Human colon cancer
14	42.6	14.5	2378	23	ABQ56733	Human prostate exp
15	42.6	14.5	2378	23	ABV23177	Human prostate exp
16	40.4	13.7	1879	23	ABV29017	L. esculentum germ
17	40.4	13.7	2024	20	AAX89865	L. esculentum germ
18	40.4	13.7	2024	21	AAA38935	Tomato germacrene
19	39.6	13.5	628	22	AAH87656	Peppermint plant o
20	36	12.2	2000	24	ABZ16713	Arabidopsis thalia
21	35.2	12.0	14041	21	AAH48024	Arabidopsis thalia
22	33.8	11.5	3744	21	AAA70149	Internal control B
23	33.8	11.5	580073	18	AAT58840	Plasmodium falcipa
24	33.6	11.4	831	25	ABX08179	Mycoplasma genital
25	33.6	11.4	2349	21	AAA05514	S. pneumoniae type
26	33.6	11.4	2393	21	AAV52223	Streptococcus pneu
27	33.2	11.3	2162598	25	ABS54454	Streptococcus pneu
28	33.2	11.3	665	22	AAL15957	Human breast cance
29	33.2	11.3	1273	24	ABZ15600	Arabidopsis thalia
30	33	11.2	3564	24	ABN66233	Streptococcus poly
31	33	11.2	4300	22	AAH42107	Nucleotide sequenc
32	33	11.2	640681	24	ABA92787	Buchnera sp. genom
33	32.8	11.2	477	21	AAC27964	Buchnera sp. genom
34	32.8	11.2	2862	24	AAI99844	Bacillus secreted pro
35	32.8	11.2	3813	24	AAI99855	Bacillus KSM-P358
36	32.8	11.2	202001	24	ABS52506	Bacillus KSM-P358
37	32.6	11.1	1728	20	AAX99301	C. elegans CED-6 p
38	32.6	11.1	2832	24	ABZ34930	Human gene express
39	32.6	11.1	2971	22	AAH54085	S. epidermidis gen
40	32.6	11.1	4029	10	AAN91323	DNA encoding rabbi
41	32.6	11.1	4029	19	AAT94064	Rabbit growth horm
42	32.4	11.0	353	23	ABV17967	Human prostate exp
43	32.4	11.0	398	23	ABV17729	Human prostate exp
44	32.4	11.0	451	23	ABV47458	Human prostate exp
45	32.4	11.0	492	23	ABV47522	Human prostate exp

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	294	100.0	294	24 AAD40074	Human IRAK4 DD (de
2	294	100.0	833	21 AAA09319	Human cancer assoc
3	294	100.0	1383	24 AAD40079	Human IRAK4 gene #
4	294	100.0	1668	23 AAS76805	DNA encoding novel
5	294	100.0	2817	24 AAD40085	Human IRAK4 gene #
6	292.4	99.5	1383	22 AAD10197	Human interleukin-
7	241.6	82.2	501	24 AAS61608	Lung small cell ca
8	204.4	69.5	1542	22 AAD10198	Mouse interleukin-

ALIGNMENTS

RESULT 1
AAD40074

ID   AAD40074 standard; cDNA; 294 BP.
XX
AC   AAD40074;
XX
DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 DD (death domain) cDNA.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   1..294
FT   /tag= a
FT   /product= "Human IRAK4 DD"
FT   /note= "No start and stop codon"
FT   /partial
XX
PN   WO200240680-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US48844.

RESULT 2

Left column

```
PR   17-NOV-2000; 200US-0715893.
PR   29-JUN-2001; 2001US-301889P.
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24854.
XX
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection and other diseases
PT   -
XX
PS   Claim 18; Page 173-174; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis , C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 DD cDNA.
XX
SQ   Sequence 294 BP; 84 A; 55 C; 62 G; 93 T; 0 other;

Query Match      100.0%;   Score 294;  DB 24;   Length 294;
Best Local Similarity   100.0%;   Pred. No. 1.6e-77;
Matches  294;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy         1 ACATATGTGGGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ACATATGTGGGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60

Qy        61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGGTGATGATAGATAC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGGTGATGATAGATAC 120

Qy       121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180

Qy       181 GAATTACTGTTTGACTGGGGGCCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GAATTACTGTTTGACTGGGGGCCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 240

Qy       241 ATCCAAAAATGAATTTTTTTGCTCCTCCCAGATGCTGTTCCC 294
             ||||||||||||||||||||||||||||||||||||||||||
Db       241 ATCCAAAAATGAATTTTTTTGCTCCTCCCAGATGCTGTTCCC 294
```

Right column

```
AAA09319
ID   AAA09319 standard; DNA; 833 BP.
XX
AC   AAA09319;
XX
DT   10-AUG-2000  (first entry)
XX
DE   Human cancer associated antigen precursor DNA, clone NY-REN-64.
XX
KW   renal cancer; cancer associated antigen precursor; diagnosis;
KW   cytostatic; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             50..670
FT                   /*tag= a
XX
PN   WO20020587-A2.
XX
PD   13-APR-2000.
XX
PF   04-OCT-1999;    99WO-US22873.
XX
PR   05-OCT-1998;    98US-0166300.
PR   05-OCT-1998;    98US-0166350.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y, Gout I, Tureci O, Sahin U, Pfreundschuh M, Scanlan MJ;
PI   Stockert E, Chen Y, Old LJ, Jager E, Knuth A;
XX
DR   WPI; 2000-303774/26.
DR   P-PSDB; AAY92347.
XX
PT   Preventing, diagnosing and/or treating disorders associated with
PT   abnormal expression of human cancer associated antigens
XX
PS   Claim 57; Page 85; 121pp; English.
XX
CC   AAA09310-20 are novel genes isolated by SEREX screening from a renal
CC   cancer cell line 1973/10.4. The genes encode cancer associated antigen
CC   precursors. These gene products are useful in methods for preventing,
CC   diagnosing and/or treating disorders, especially cancer, associated with
CC   abnormal expression of human cancer associated antigens. The method
CC   comprises contacting a sample from a subject with an agent that
CC   specifically binds to the nucleic acid molecule or expression product
CC   (or fragment) complexed with a human leukocyte antigen (HLA) molecule
CC   and determining the interaction between the agent and the nucleic acid
CC   molecule or the expression product as a determination of the disorder.
XX
SQ   Sequence 833 BP; 273 A; 155 C; 179 G; 226 T; 0 other;

Query Match      100.0%;   Score 294;  DB 21;   Length 833;
Best Local Similarity   100.0%;   Pred. No. 2.2e-77;
Matches  294;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy         1 ACATATGTGGGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        74 ACATATGTGGGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 133

Qy        61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGGTGATGATAGATAC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGGTGATGATAGATAC 193

Qy       121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       194 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 253

Qy       181 GAATTACTGTTTGACTGGGGGCCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       254 GAATTACTGTTTGACTGGGGGCCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 313
```

```
Qy   241  ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCGTGTTCCC  294
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314  ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCGTGTTCCC  367
```

RESULT 3

ID   AAD40079 standard; DNA; 1383 BP.
AC   AAD40079;
DT   22-OCT-2002 (first entry)
DE   Human IRAK4 gene #1.

KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.

OS   Homo sapiens.

FH   Key   Location/Qualifiers
FT   CDS   1..1383
FT   /*tag= a
FT   /product= "Human IRAK4"

PN   WO200240680-A2.

PD   23-MAY-2002.

PF   15-NOV-2000; 2001WO-US44844.

PR   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.

PA   (BURN-) BURNHAM INST.

PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;

DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24859.

PT   The invention relates to an isolated polypeptide comprising a death
PT   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
PT   is useful for identifying a binding agent, preferably a protein or a drug
PT   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
PT   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
PT   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
PT   detecting the association of the domain and the candidate binding agent
PT   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
PT   chemical crosslinking, nuclear magnetic resonance (NMR), mass
PT   spectroscopy (MS) and FPA. The invention is useful for modulating the
PT   level of a cell process such as cell proliferation, cell adhesion, cell
PT   stress responses, responses to microbial infection and B cell
PT   immunoglobulin class switching, in particular apoptosis within a cell.
PT   Antibody specifically reactive with CTDD DD of C. trachomatis, C.
PT   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
PT   CTDD DD protein is useful for detecting a Chlamydia infection. The
PT   invention is useful for modulating the activity of oncogenic proteins,
PT   for treating a pathology caused by the oncogenic activity of bacterial
PT   proteins. The protein and antibody specific for it are useful for
PT   treating bacterial infections by modulating the activity of bacterial
PT   proteins. The protein is useful for discovery of drugs that suppress
PT   -

PS   Claim 19; Page 180-182; 209pp; English.

CC   New polypeptide comprising a death domain or death effector domain,
CC   useful for discovery of drugs that suppress infection, inflammation,
CC   allergy, sepsis, autoimmunity, allograft rejection and other diseases

CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathology, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 gene.

SQ   Sequence 1383 BP; 463 A; 243 C; 283 G; 394 T; 0 other;

Query Match   100.0%; Score 294; DB 24; Length 1383;
Best Local Similarity   100.0%;   Pred. No. 2.6e-77;
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1  ACATATATGTGCGCTGCCCTCAATGTTGGACTAATTAGGAGAGCTGTCAGATTTATTGATCCT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    25  ACATATATGTGCGCTGCCCTCAATGTTGGACTAATTAGGAGAGCTGTCAGATTTATTGATCCT   84

Qy    61  CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    85  CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC  144

Qy   121  AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145  AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT  204

Qy   181  GAATTACTTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGTGATCTTCTTGTTG  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205  GAATTACTTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGTGATCTTCTTGTTG  264

Qy   241  ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCGTGTTCCC  294
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   265  ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCGTGTTCCC  318
```

RESULT 4

ID   AAS76805 standard; cDNA; 1668 BP.
AC   AAS76805;
DT   13-FEB-2002 (first entry)
DE   DNA encoding novel human diagnostic protein #12609.

KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS   Homo sapiens.

PN   WO200175067-A2.

PD   11-OCT-2001.

PF   30-MAR-2001; 2001WO-US08631.

PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.

PA   (HYSE-) HYSEQ INC.

PI   Drmanac RT, Liu C, Tang YT;

DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12618.

PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostic, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -

PS   Claim 1; SEQ ID No 12609; 103pp; English.

CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques,
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement.(II) and its binding partners are useful in medical
CC   imaging of sites expressing (II).(I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1668 BP; 289 C; 332 G; 476 T; 0 other;

Query Match   100.0%;
Best Local Similarity   100.0%;
Conservative   100.0%; Score 294; DB 23; Length 1668;
Matches 294;   0; Pred. No. 2.8e-77;
                Mismatches   0; Indels   0; Gaps   0;

Qy   1 ACATATATGTGCGGCTCGCCTCAATGTTGGACTAATTAGGAAGCGTGTCAGATTTTATTGATCCT 60
        ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACATAT GTGCGGCTCGCCTCAATGTTGGACTAATTAGGAAGCGTGTCAGATTTTATTGATCCT 84

Qy   61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTTATTAAAAAACCATCTGGTGATGATGATAGATAC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTTATTAAAAAACCATCTGGTGATGATGATAGATAC 144

Qy 121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 145 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 204

Qy 181 GAATTACTGTTTGACTGGGGCGCCACACAAAATTGCACAGTTGGTGATCTTGTGGGATCTTTTG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 205 GAATTACTGTTTGACTGGGGCGCCACACAAAATTGCACAGTTGGTGATCTTGTGGGATCTTTTG 264

Qy 241 ATCCAAAATGAAATTTTTTGCTCCTCGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 265 ATCCAAAATGAAATTTTTTGCTCCTCGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 318

RESULT 5
AAD40085
ID   AAD40085 standard; DNA; 2817 BP.
XX
AC   AAD40085;
XX
DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 gene #2.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   50..1432
FT   /tag= a
FT   /product= "Human IRAK4"
XX
PN   WO200240680-A2.
XX
PD   23-MAY-2002.

XX
PF   15-NOV-2001; 2001WO-US44844.
XX
PF   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.
XX
PA   (BURN-) BURNHAM INST.
XX
PA   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24865.
XX
CC   New polypeptide comprising a death domain or death effector domain,
CC   useful for discovery of drugs that suppress infection, inflammation,
CC   allergy, sepsis, autoimmunity, allograft rejection and other diseases
PT   -
XX
PS   Claim 19; Page 194-196; 209pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the domain and the candidate binding agent
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the
CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis, C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 gene.
XX
SQ   Sequence 2817 BP; 912 A; 547 C; 586 G; 772 T; 0 other;

Query Match   100.0%;
Best Local Similarity   100.0%;
Conservative   100.0%; Score 294; DB 24; Length 2817;
Matches 294;   0; Pred. No. 3.3e-77;
                Mismatches   0; Indels   0; Gaps   0;

Qy   1 ACATATATGTGCGGCTCGCCTCAATGTTGGACTAATTAGGAAGCGTGTCAGATTTTATTGATCCT 60
        ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74 ACATAT GTGCGGCTCGCCTCAATGTTGGACTAATTAGGAAGCGTGTCAGATTTTATTGATCCT 133

Qy   61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTTATTAAAAAACCATCTGGTGATGATGATAGATAC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTTATTAAAAAACCATCTGGTGATGATGATAGATAC 193

Qy 121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 194 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 253

Qy 181 GAATTACTGTTTGACTGGGGCGCCACACAAAATTGCACAGTTGGTGATCTTGTGGGATCTTTTG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 254 GAATTACTGTTTGACTGGGGCGCCACACAAAATTGCACAGTTGGTGATCTTGTGGGATCTTTTG 313

Qy 241 ATCCAAAATGAAATTTTTTGCTCCTCGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 314 ATCCAAAATGAAATTTTTTGCTCCTCGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 367

RESULT 6

ID   AAD10197 standard; cDNA; 1383 BP.
XX
AC   AAD10197;
XX
DT   24-SEP-2001 (first entry)
XX
DE   Human interleukin-1 (IL-1) receptor-associated kinase (IRAK)-4 cDNA.
XX
KW   Human; interleukin-1 receptor associated kinase-4; IRAK-4; cytostatic;
KW   IL; antibacterial; antiinflammatory; ophthalmological; vasotropic; OPD;
KW   chronic obstructive pulmonary disease; neuroprotective; chronic cough;
KW   adult respiratory disease syndrome; pulmonary fibrosis; asthma; ARDS;
KW   interstitial lung disease; allergic rhinitis; transplant rejection;
KW   autoimmune disease; rheumatoid arthritis; systemic lupus erythematosus;
KW   multiple sclerosis; diabetes; cancer; solid tumour; lymphoma; stroke;
KW   cardiovascular disease; atherosclerosis; neurodegenerative disease;
KW   sepsis; osteoarthritis; rash; osteoporosis; psoriasis; dermatitis;
KW   inflammatory bowel disease; gout; Crohn's disease; ulcerative colitis;
KW   Behcet's syndrome; ankylosing spondylitis; IL-1 receptor/Toll receptor;
KW   sarcoidosis; transgenic animal.; SS.
XX
OS   Homo sapiens.
XX
FH
FT   Key   Location/Qualifiers
FT   CDS   1..1383
FT   /*tag= a
FT   /product= "Human IRAK-4"
XX
PN   WO200151641-A1.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US01171.
XX
PR   13-JAN-2000; 2000US-0176395.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Wesche H, Li S;
XX
DR   WPI; 2001-451860/48.
DR   P-PSDB; AAE05398.
XX
PT   Novel human interleukin-1 receptor associated kinase polypeptide,
PT   useful for identifying modulators of the polypeptide for treating gout,
PT   asthma, allergic rhinitis, multiple sclerosis and skin cancer -
XX
PS   Claim 7; Fig 2; 89pp; English.
XX
CC   The present sequence is a cDNA encoding human interleukin (IL)-1 receptor
CC   associated kinase (IRAK)-4. IRAK associate with activated IL-1, IL-18
CC   and other receptors and act to transduce signals originating from the
CC   activated receptors, ultimately leading to a variety of downstream
CC   effects such as nuclear factor (NF)-kappaB activation. The IRAK-4
CC   inhibitors are useful for treating inflammatory diseases such as
CC   pulmonary diseases and diseases of the airway (e.g., adult respiratory
CC   disease syndrome (ARDS), chronic obstructive pulmonary disease (OPD),
CC   pulmonary fibrosis, interstitial lung disease, chronic cough
CC   or allergic rhinitis), transplant rejection, autoimmune diseases (e.g.,
CC   rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis
CC   or diabetes), cancer (e.g., solid tumours, skin cancer or lymphoma),
CC   cardiovascular diseases (e.g., stroke and atherosclerosis), diseases
CC   of the central nervous system (e.g., neurodegenerative disease), diseases
CC   mediated sepsis, non-CD14 mediated sepsis, osteoarthritis, osteoporosis,
CC   psoriasis, diseases of the skin (e.g., rash, contact dermatitis, atopic
CC   dermatitis), inflammatory bowel disease (e.g., Crohn's disease and
CC   ulcerative colitis), Behcet's syndrome, ankylosing spondylitis, gout,
CC   sarcoidosis and ophthalmic diseases and conditions. The inhibitors of
CC   IRAK-4 activity or expression are used to inhibit signal transduction
CC   resulting from the activation of an interleukin-1 receptor (IL-1R)/Toll
CC   receptor in a cell. They also inhibit the activation of a transcription
CC   factor that activates NFkappaB in the cell. IRAK-4 is used to create a
CC   nonhuman transgenic animal which is useful for testing the function of
CC   IRAK-4 in vivo, to generate models for the study of potential treatments
CC   disorders and conditions and for the development of potential treatments
CC   for IRAK-4-related inflammatory diseases and conditions. IRAK-4 sequences
CC   are also used in gene therapy and in antisense therapy.
XX
SQ   Sequence 1383 BP; 463 A; 244 C; 283 G; 393 T; 0 other;

Query Match   99.5%; Score 292.4; DB 22; Length 1383;
Best Local Similarity   99.7%; Pred. No. 7.8e-77;
Matches 293; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 ACATATGTGCCCTGCCTCAATGTTGGACTAATTAGGAAAGCTGTCAGATTTTATTGATCCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACATATGTGCCCTGCCTCAATGTTGGACTAATTAGGAAAGCTGTCAGATTTTATTGATCCT 84

Qy   61 CAAGAAGGATGGAAGAGAGTTAGCTGTAGCTATTAAAAAACCATTCTGGTGATGATAGATAC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAAGAAGGATGGAAGAGAAGTTAGCTGTAGCTATTAAAAAACCATTCTGGTGATGATAGATAC 144

Qy   121 AATCAGTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGGAAAAGTCCCACTTCT 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145 AATCAGTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGGAAAAAGTCCCACTTCT 204

Qy   181 GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGCTGGTGATCTTGTGGATCTTTTG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205 GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGCTGGTGATCTTGTGGATCTTTTG 264

Qy   241 ATCCAAAATGAATTTTTTGCTCCTGGGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   265 ATCCAAAATGAATTTTTTGCTCCTGGGAGTCTTTTGCTCCCAGATGCTGTTCCC 318

RESULT 7

ID   AAS61608 standard; cDNA; 501 BP.
XX
AC   AAS61608;
XX
DT   29-JAN-2002 (first entry)
XX
DE   Lung small cell carcinoma antigen, cDNA #149.
XX
KW   Human; cytostatic; antitumour; lung small cell cancer antigen;
KW   tumour; lung cancer; SS.
XX
OS   Homo sapiens.
XX
PN   WO200177168-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-US11859.
XX
PR   11-APR-2000; 2000US-196780P.
PR   21-JUN-2000; 2000US-213361P.
PR   01-SEP-2000; 2000US-229763P.
PR   05-SEP-2000; 2000US-230629P.
PR   14-SEP-2000; 2000US-232565P.
PR   19-DEC-2000; 2000US-257037P.
PR   08-JAN-2001; 2001US-260796P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Lodes MJ, Wang T, Mohamath R, Indirias CY;
XX
DR   WPI; 2002-010896/01.
XX
PT   Lung tumour polynucleotide and polypeptides useful in therapy and
PT   diagnosis of cancer especially lung cancer -

RESULT 8

AAD10198

ID   AAD10198 standard; cDNA; 1542 BP.
XX
AC   AAD10198;
XX
DT   24-SEP-2001 (first entry)
XX
DE   Mouse interleukin-1 (IL-1) receptor-associated kinase (IRAK)-4 cDNA.
XX
CC   Mouse; interleukin-1 receptor associated kinase-4; IRAK-4; cytostatic;
CC   IL; antibacterial; antiinflammatory; ophthalmological; vasotropic; OPD;
CC   chronic obstructive pulmonary disease; neuroprotective; chronic cough;
CC   adult respiratory disease syndrome; pulmonary fibrosis; asthma; ARDS;
CC   interstitial lung disease; allergic rhinitis; transplant rejection;
CC   autoimmune disease; rheumatoid arthritis; systemic lupus erythematosus;
CC   multiple sclerosis; diabetes; cancer; solid tumour; lymphoma; stroke;
CC   cardiovascular disease; atherosclerosis; neurodegenerative disease;
CC   sepsis; osteoarthritis; rash; osteoporosis; psoriasis; dermatitis;
CC   inflammatory bowel disease; gout; Crohn's disease; ulcerative colitis;
CC   Behcet's syndrome; ankylosing spondylitis; IL-1 receptor/Toll receptor;
CC   sarcoidosis; transgenic animal; ss.
XX
OS   Mus sp.
XX
XX   Key   Location/Qualifiers
     CDS   1..1542
                      /*tag= a
                      /product= "Mouse IRAK-4"
XX
PN   WO20015164l-A1.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US01171.
XX
PR   13-JAN-2000; 2000US-0176395.
XX
PA   (TULA-) TULARIK INC.
XX
PI   Wesche H, Li S;
XX
DR   WPI; 2001-451860/48.
DR   P-PSDB; AAE05399.
XX
PT   Novel human interleukin-1 receptor associated kinase polypeptide,
PT   useful for identifying modulators of the polypeptide for treating gout,
PT   asthma, allergic rhinitis, multiple sclerosis and skin cancer -

Claim 26; Fig 4; 89pp; English.

XX
PS
XX
CC   The present sequence is a cDNA encoding mouse interleukin (IL)-1 receptor
CC   associated kinase (IRAK)-4. IRAK associate with activated IL-1, IL-18
CC   and other receptors and act to transduce signals originating from the
CC   activated receptor, ultimately leading to a variety of downstream
CC   effects such as nuclear factor (NF)-kappaB activation. The IRAK-4
CC   inhibitors are useful for treating inflammatory diseases such as
CC   pulmonary diseases and diseases of the airway (e.g., adult respiratory
CC   disease syndrome (ARDS), chronic obstructive pulmonary disease (OPD),
CC   pulmonary fibrosis, interstitial lung disease, chronic cough
CC   or allergic rhinitis), transplant rejection, autoimmune diseases (e.g.,
CC   rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis
CC   or diabetes), cancer (e.g., solid tumours, skin cancer or lymphoma),
CC   cardiovascular diseases (e.g., stroke and atherosclerosis), diseases
CC   of the central nervous system (e.g., neurodegenerative disease), CD14
CC   mediated sepsis, non-CD14 mediated sepsis, osteoarthritis, osteoporosis,
CC   psoriasis, diseases of the skin (e.g., rash, contact dermatitis, atopic
CC   dermatitis), inflammatory bowel disease (e.g., Crohn's disease and
CC   ulcerative colitis), Behcet's syndrome, ankylosing spondylitis, gout,
CC   sarcoidosis and ophthalmic diseases and conditions. The inhibitors of
CC   IRAK-4 activity or expression are used to inhibit signal transduction
CC   resulting from the activation of an interleukin-1 receptor (IL-IR)/Toll
CC   receptor in a cell. They also inhibit the activation of a transcription
CC   factor that activates NFkappaB in the cell. IRAK-4 is used to create a
CC   nonhuman transgenic animal which is useful for testing the function of
CC   IRAK-4 in vivo, to generate models for the study of inflammatory
CC   disorders and conditions and for the development of potential treatments
CC   for IRAK-4-related inflammatory diseases and conditions. IRAK-4 sequences
CC   are also used in gene therapy and in antisense therapy.
XX
SQ   Sequence 1542 BP; 421 A; 392 C; 423 G; 306 T; 0 other;

Query Match   69.5%; Score 204.4; DB 22; Length 1542;
Best Local Similarity   81.0%; Pred. No. 1.1e-50;
Matches 238; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy   1 ACATATGTGCGCCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCTT 60
         ||||| || ||| | |||| |||| || | |||||||||||| |||||||||||||||
Db 187 ACATACCATACGCAACCTTAATGTGGGGATCCTTAGGAAGCTGTCGGATTTTATTGATCCT 246

Qy   61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
         |||||||| |||||||| |||| || ||||| |||||| || || || || || ||||||
Db 247 CAAGAAGGGTGGAAGAAATTAGCAGTGGCTATCAAAAAGCCGTCCGGCGACGACAGATAC 306

Qy 121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
         |||||||| || ||||||||| || |||| |||||||| ||| |||| | ||||| ||
Db 307 AATCAGTTCCATATAAGGAGGATTCGAAGCCTTACTTCAGACCGGCAGGAAGAGCCCCACCTGT 366

Qy   181 GAATTACTGTTTGACTGGGGCCACCAAAATTGCACAGTTGGTGATCTTTGTGATCTTTTG 240
          ||| ||||||| |||||||||||||||||||||||||||||| ||||||||||||
Db   367 GAACTGCTGTTTGACTGGGGCCACCACGAACTGCACAGTTGGCGACCTTGTGGATCTACTG 426

Qy   241 ATCCAAAATGAATTTTTTGCTCCTGGGAGTCTTTGCCTCCAGATGCTGTTTCCC 294
          || ||||||||||||||||||||||| |||||| || |||||||||||||||||
Db   427 GTCCAGATTGAGCTGTTTGCCCCGCCAACTCCTCCTGCGCCCGGATGGCCGTTCCC 480

RESULT 9
ID   AAS76803 standard; cDNA; 405 BP.
XX
AC   AAS76803;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12607.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12616.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 12607; 103pp; English.

CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (I) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.

SQ   Sequence 405 BP; 93 A; 87 C; 83 G; 142 T; 0 other;

Query Match   51.2%; Score 150.4; DB 23; Length 405;

Best Local Similarity 99.3%; Pred. No. 7.9e-35; Mismatches 1; Indels 0; Gaps 0;
Matches 151; Conservative 0;

Qy   134 TAAGGAGANTTGAAGCANTACTTACTTCAAAACTGGAAAAAAGTCCCACTTCTGAATTAC 193
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   115 TAAGGAGANTTGAAGCANTACTTACTTCAAAACTGGAAAAAAGTCCCACTTCTGAATTAC 174

Qy   194 ACTGGCGGCCACCACAAAATTGCACAGTTGGTGATCTTTGTGGGATCTTTTGATCCAAAATGAAT 253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   175 ACTGGCGGCCACCACAAAATTGCACAGTTGGTGATCTTTGTGGGATCTTTTGATCCAAAATGAAT 234

Qy   254 TTTTTGCTCCTGGGAGTCTTTTTGCTCCCAGAT 285
          |||||||||||||||||||||||||||||||||
Db   235 TTTTTGCTCCTGGGAGTCTTTTTGCTCCCAGGT 266

RESULT 10
ID   AAD40084 standard; DNA; 211 BP.
XX
AC   AAD40084;
XX
DT   22-OCT-2002 (first entry)
XX
DE   Human IRAK4 short gene.
XX
KW   Human; death domain; DD; death effector domain; DED; Chlamydia infection;
KW   NB-ARC domain; apoptosis; oncogenic protein; bacterial infection; sepsis;
KW   inflammation; allergy; autoimmunity; allograft rejection; cell division;
KW   immune-based pathology; fibrosis; arthritis; graft versus host disease;
KW   immunosuppressive; gene therapy; antisense therapy; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200240680-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US44844.
XX
PR   17-NOV-2000; 2000US-0715893.
PR   29-JUN-2001; 2001US-301889P.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..180
FT   /*tag= a
FT   /product= "Human IRAK4 short protein"
XX
PA   (BURN-) BURNHAM INST.
XX
PI   Reed JC, Godzik A, Pawlowski K, Fiorentino L, Lee SH, Roth W;
PI   Stenner-liewen F;
XX
DR   WPI; 2002-500222/53.
DR   P-PSDB; AAE24864.
XX
PT   New polypeptide comprising a death domain or death effector domain,
PT   useful for discovery of drugs that suppress infection, inflammation,
PT   allergy, sepsis, autoimmunity, allograft rejection and other diseases
-
XX
PS   Claim 19; Page 195; 209pp; English.

CC   The invention relates to an isolated polypeptide comprising a death
CC   domain (DD), death effector domain (DED) or NB-ARC domain. The invention
CC   is useful for identifying a binding agent, preferably a protein or a drug
CC   that binds a DD, DED or NB-ARC domain, by contacting a DD, DED or NB-ARC
CC   domain from DAP3, IRAK4, CTDD (Chlamydia trachomatis DD protein), DED4 or
CC   NIDD (NGFR-interacting Death Domain), with a candidate binding agent and
CC   detecting the association of the candidate binding agent,
CC   by yeast two hybrid assay, immunoprecipitation, SPA, ultraviolet (UV) or
CC   chemical crosslinking, nuclear magnetic resonance (NMR), mass
CC   spectroscopy (MS) and FPA. The invention is useful for modulating the

CC   level of a cell process such as cell proliferation, cell adhesion, cell
CC   stress responses, responses to microbial infection and B cell
CC   immunoglobulin class switching, in particular apoptosis within a cell.
CC   Antibody specifically reactive with CTDD DD of C. trachomatis, C.
CC   muridarum, C. pneumoniae, and C. psittaci or a nucleic acid encoding the
CC   CTDD DD protein is useful for detecting a Chlamydia infection. The
CC   invention is useful for modulating the activity of oncogenic proteins,
CC   for treating a pathology caused by the oncogenic proteins and for
CC   treating bacterial infections by modulating the activity of bacterial
CC   proteins. The protein and antibody specific for it are useful for
CC   discovery of drugs that suppress infection, inflammation, allergy,
CC   sepsis, autoimmunity, allograft rejection and other diseases. The protein
CC   is useful for treating immune-based pathologies, pathologies associated
CC   with cell division, inflammatory diseases such as sepsis, fibrosis,
CC   arthritis, graft versus host disease. The invention is used in antisense
CC   therapy and gene therapy. The present sequence is human IRAK4 short gene.
XX
SQ   Sequence 211 BP; 77 A; 41 C; 36 G; 57 T; 0 other;

Query Match   46.8%; Score 137.6; DB 24; Length 211;
Best Local Similarity   97.2%; Pred. No.4.1e-31;
Matches 140;   Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   1 ACATATGTGCGCGCTGCCTCAATGTTGGACACTAATTAGGAAGCTGTCAGAGATTTATTGATCCT 60
              ||||||||||||| ||||||||||||||||| |||||||| |||||||||| ||||||||||||
Db   25 ACATATGTGCGGCTGCCTCAATGTTGGACTAATTAGGACGCTGTCGATTTTATTGATCCT 84

Qy   61 CAAGAAGGATGGAAGAAGTTAGCGTAGCTGTATTAAAAACCATCTGGTGATGADAGATAC 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTGTATTAAAAACCATCTGGTGATGADAGATAC 144

Qy   121 AATCAGTTTCACATAAGGAGAGATTT 144
              |||||||||||||||| ||||||
Db   145 AATCAGTTTCACATAAGATGCTGT 168

RESULT 11
AAS76802
ID   AAS76802 standard; cDNA; 415 BP.
AC   AAS76802;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #12606.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG12615.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 12606; 103pp; English.

XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 415 BP; 144 A; 83 C; 78 G; 110 T; 0 other;

Query Match   46.8%; Score 137.6; DB 23; Length 415;
Best Local Similarity   97.2%; Pred. No. 5.1e-31;
Matches 140;   Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   1 ACATATGTGCGCGCTGCCTCAATGTTGGACACTAATTAGGAAGCTGTCAGAGATTTATTGATCCT 60
              ||||||||||||| ||||||||||||||||| |||||||| |||||||||| ||||||||||||
Db   60 ACATATGTGCGGCTGCCTCAATGTTGGACTAATTAGGACGCTGTCGATTTTATTGATCCT 119

Qy   61 CAAGAAGGATGGAAGAAGTTAGCGTAGCTGTATTAAAAACCATCTGGTGATGADAGATAC 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   120 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTGTATTAAAAACCATCTGGTGATGADAGATAC 179

Qy   121 AATCAGTTTCACATAAGGAGAGATTT 144
              |||||||||||||||| ||||||
Db   180 AATCAGTTTCACATAAGATGCTGT 203

RESULT 12
AAH13798
ID   AAH13798 standard; cDNA; 2213 BP.
AC   AAH13798;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:10742.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
PR   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 10742; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are also useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AaHO3166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotide, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2213 BP; 642 A; 488 C; 516 G; 567 T; 0 other;

Query Match   46.8%; Score 137.6; DB 22; Length 2213;
Best Local Similarity 97.2%; Pred. No. 8.9e-31;
Matches 140; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy    1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAGAAGCTGTCAGATTTTATTGATCCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  775 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAGAAGCTGTCAGATTTTATTGATCCT 834

Qy   61 CAAGAAGGATGGAAGAAGAGTTAGCTGTGTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  835 CAAGAAGGATGGAAGAAGAGTTAGCTGTGTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 894

Qy  121 AATCAGTTTCACATAAGGAGATTT 144
          ||||||||||||||||||||||||
Db  895 AATCAGTTTCACATAAGGAGATGCTGT 918
```

RESULT 13
ABQ56733
ID   ABQ56733 standard; cDNA; 408 BP.
XX
AC   ABQ56733;
XX
DT   02-AUG-2002 (first entry)
XX
DE   Human colon cancer related nucleotide sequence SEQ ID NO:428.
XX
KW   Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW   genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20029086-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30732.
XX
PR   02-OCT-2000; 2000US-237271P.

XX
PA   (FARB) BAYER CORP.
XX
PI   Burgess C, Astle JH, Carroll B, Catino TJ, Dwivedi P, Molino GA;
PI   Thiagilngam A, Lewis ME;
XX
DR   WPI; 2002-426115/45.
XX
PS   Claim 1; Fig 1; 796pp; English.
XX
PT   New isolated nucleic acid that is differentially expressed in cancer
PT   tissues useful for determining the presence of colon cancer in a cell
PT   or tissue type, and in antisense therapy -
XX
CC   ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC   expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
CC   encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC   used in antisense therapy. An antibody immunoreactive with a polypeptide
CC   encoded by (I) is useful for detecting cancer in a patient sample, and
CC   for detecting the presence or absence of a polynucleotide encoded by a
CC   nucleic acid which hybridises to (I) in a cell. A probe/primer derived
CC   from (I) can be used for determining the presence of a nucleic acid which
CC   hybridises to (I), and for determining the phenotype of cells in a sample
CC   of cells from a patient. (I) is useful for determining the presence of
CC   colon cancer in a cell or tissue type, for determining the presence or
CC   state of other type of cancer, in antisense therapy, to generate
CC   macroarrays on a solid surface, to identify a chromosome on which the
CC   corresponding gene resides, and in tissue profiling, forensics, genetic
CC   analysis, mapping and diagnostic applications. (I) can be used to raise
CC   antibodies, and to screen for peptide analogues and antagonists.
XX
SQ   Sequence 408 BP; 114 A; 92 C; 81 G; 61 T; 60 other;

Query Match   35.6%; Score 104.8; DB 24; Length 408;
Best Local Similarity 69.5%; Pred. No. 2.8e-21;
Matches 207; Conservative 0; Mismatches 81; Indels 10; Gaps 7;

```
Qy    1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAGAAGCTGTCAGATTTTATTGATCCT 60
          |||||||||||||| ||||   ||||| ||||||||||||||||||||||||||||||||
Db   39 ACNTGTCGGGNCCCTGCCNNATGTGGCNNCTAATNANAGCNGNCAGANTTTANCGNANCCT 98

Qy   61 CAAGAAGGATGGAAGAAGAGTTAGCTGTGTAGCTATTAAAAAAACCATCTGGTGATGAT-AGATA 119
          ||||||  |||| ||||||||||||||||| |||||||||||||||||||||| ||||| |||||
Db   99 NNANANGANGG-ANAAGGCAGCTGNAGCTATNAAAAAAACCANCNGGGATGATNNGATA 157

Qy  120 CAATCAGTTTCACATA-AGGAGATTTGAAGCATTACTTCAAA---CTGGAAAAAGTCCCA 175
          |||||||||||||||| ||| |||||||  ||||| ||| || |||||| ||||||||||
Db  158 CCATCAGTTTCACANANNGGAGANNTCAAGCANTACTNNAAAAACTGGGAAAAAGGCCN 217

Qy  176 CTTCTGAAT-TACTGTTTGACTGGGGGGCACCACAAATTGCACAGTT-GGTGATCTTGTGGA 233
          ||||| ||| ||||||||| |||||||||||||||||||||||||| |||||||||||||
Db  218 CTTCGGAAANGAACTGTTTGACTGGGGCCACCACAAAATTGCACAGNTGGGGGGATCTTGTGNA 277

Qy  234 TCTTTTGATCCAAAAATGAATTTTTTTG--CTCCTGCGAGTCTTTTTGCTCCCACAGATGCTG 289
          |||| |  |||||||||||||||||||   |||  || ||| ||||||  || ||||||| ||
Db  278 TNCCNCGAANCACAACNGAAATTNTNGCTCCNTGCNANGCCNCGGGGCNCCCAGACGCNG 335
```

RESULT 14
ABV23177
ID   ABV23177 standard; cDNA; 2378 BP.
XX
AC   ABV23177;
XX
DT   16-SEP-2002 (first entry)
XX
DE   Human prostate expression marker cDNA 23168.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX

PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R, Endege WO, Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS   Claim 1; Page 4160-4161; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (II) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 2378 BP; 769 A; 396 C; 487 G; 706 T; 20 other;

Query Match   14.5%; Score 42.6; DB 23; Length 2378;
Best Local Similarity   49.8%; Pred. No. 0.015;
Matches   108; Conservative   0; Mismatches 109; Indels   0; Gaps   0;

Qy   59 CTCAAGAAGGAATGGAGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATGAT 118
               ||| | |||
Db   467 CACACGAAAATTTAGAAAGCTGCTTTCTAAAGAACCTAATCCACCAATAGATCAAGTTATAC 526

Qy   119 ACAATCAGTTTCACATAAGAGGAGATTTGAAGCCATTACTTCAAACTGGAAAAAGTCCACTT 178
               | ||| |
Db   527 AGAAACCAGGAGTTGTACAGAGATTTGTGAAGAATTTCTTGAAGAAATGAAAATTGCACTT 586

Qy   179 CTGAATTACTGTTTGACTGGGGGCACCACAAAATTGCACAGTTGGGTGATCTTGTGGATCTTT 238
               | ||| | |||
Db   587 TACAATTTGAAGCTGCATGGGCATTAACAAATAATAGCATCTGGACTTTTCTGCATACCA 646

Qy   239 TGATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTT 275
                | ||
Db   647 AGGTAGTGATTGAAACTGGGGCTGTTCCGATTTTTAT 683

RESULT 15
ABV29017
ID   ABV29017 standard; cDNA; 2378 BP.
XX
AC   ABV29017;
XX
DT   16-SEP-2002 (first entry)
XX
DE   Human prostate expression marker cDNA 29008.

Search completed: January 17, 2004, 23:50:44
Job time : 195 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: January 18, 2004, 00:39:27 ; Search time 214 Seconds
                                              (without alignments)
                                              4842.432 Million cell updates/sec

Title: US-10-001-254-5
Perfect score: 294
Sequence: 1 acatatgtgcgctgcctcaa.............tgctcccagatgctgtccc 294

Scoring table: IDENTITY NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Post-processing: Minimum Match 0%
Maximum DB seq length: 200000000
Minimum DB seq length: 0
Maximum Match 100%
Listing first 45 summaries

Database :   Published Applications NA:*
                 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 8: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                 9: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
                13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                15: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB ID	Description
1	294	100.0	294	15	US-10-001-254-5	Sequence 5, Appli
2	294	100.0	1383	15	US-10-001-254-15	Sequence 15, Appli
3	294	100.0	2817	11	US-09-966-451-3	Sequence 3, Appli
4	294	100.0	2817	11	US-09-938-842A-3405	Sequence 727, App
5	292.4	99.5	1383	11	US-09-759-595-2	Sequence 2, Appli
6	241.6	82.2	501	9	US-09-833-790-149	Sequence 149, App
7	204.4	69.5	1542	11	US-09-759-595-4	Sequence 4, Appli
8	151.8	51.6	31000	11	US-09-966-451-10	Sequence 10, Appli
9	137.6	46.8	211	15	US-10-001-254-25	Sequence 25, Appli
10	40.4	13.7	2024	9	US-09-887-586A-51	Sequence 51, Appli
11	40.4	13.7	2024	9	US-09-903-012-51	Sequence 51, Appli
12	40.4	13.7	2024	11	US-09-900-797-51	Sequence 51, Appli
13	36	12.2	2000	10	US-09-938-842A-4518	Sequence 4518, Ap
14	33.8	11.5	22684	13	US-09-960-858-2	Sequence 2, Appli
15	33.8	11.5	22684	13	US-09-960-870-2	Sequence 2, Appli

	16	33.8	11.5	580073	13	US-10-205-220-1	Sequence 1, Appli
	17	33.6	11.4	2349	11	US-09-769-787-287	Sequence 287, App
	18	33.2	11.3	899	15	US-10-198-846-7068	Sequence 7068, Ap
	19	33.2	11.3	1273	10	US-09-938-842A-3405	Sequence 3405, Ap
c	20	33.2	11.3	2007	10	US-09-887-576-288	Sequence 288, App
c	21	33	11.2	2345	13	US-10-027-632-110060	Sequence 110060,
c	22	33	11.2	2345	13	US-10-027-632-110061	Sequence 110061,
c	23	33	11.2	2345	14	US-10-027-632-110060	Sequence 110060,
c	24	33	11.2	2345	14	US-10-027-632-110061	Sequence 110061,
c	25	33	11.2	458	14	US-10-017-161-727	Sequence 727, App
c	26	33	11.2	14368	14	US-10-292-798-637	Sequence 637, App
c	27	33	11.2	640681	10	US-09-734-674-3	Sequence 3, Appli
c	28	32.8	11.2	458	11	US-09-790-988-1	Sequence 1, Appli
c	29	32.8	11.2	458	14	US-10-027-632-291958	Sequence 291958,
c	30	32.8	11.2	202001	9	US-09-734-674-3	Sequence 3, Appli
c	31	32.6	11.1	202001	15	US-10-027-632-282173	Sequence 42, Appli
c	32	32.2	11.0	2832	13	US-10-101-510-42	Sequence 42, Appli
c	33	32.2	11.0	658	13	US-10-027-632-282173	Sequence 282173,
c	34	32.2	11.0	658	14	US-10-027-632-282173	Sequence 282173,
c	35	32.2	11.0	895	13	US-10-027-632-121917	Sequence 121917,
c	36	32.2	11.0	895	14	US-10-027-632-121917	Sequence 121917,
c	37	32.2	11.0	2000	10	US-09-938-842A-4674	Sequence 4674, Ap
c	38	32.2	11.0	4110	15	US-10-198-846-9909	Sequence 9909, Ap
c	39	32	10.9	3673778	13	US-10-312-841-2	Sequence 2, Appli
c	40	32	10.9	275	10	US-09-878-574-15709	Sequence 15709, A
c	41	32	10.9	424	15	US-10-102-524-227	Sequence 227, Ap
c	42	32	10.9	1843	10	US-09-880-107-3412	Sequence 3412, Ap
c	43	32	10.9	3012	11	US-09-919-039-167	Sequence 167, App
c	44	32	10.9	9834	8	US-08-781-986A-37	Sequence 37, Appl
c	45	31.8	10.8	418	11	US-09-918-995-3050	Sequence 3050, Ap

ALIGNMENTS

RESULT 1
US-10-001-254-5

Sequence 5, Application US/10001254
; Publication No. US20030497021A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sug Hyung
; APPLICANT: Roth, Wilfried
; APPLICANT: Stenner-Liewen, Frank
; TITLE OF INVENTION: No. US20030497021A1e1 Death Domain Proteins
; FILE REFERENCE: P-LJ 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(294)
US-10-001-254-5

Query Match   100.0%; Score 294; DB 15; Length 294;
Best Local Similarity   100.0%;
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60

Column 1

RESULT 2
US-10-001-254-15
; Sequence 15, Application US/10001254
; Publication No. US20030049702A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sug Hyung
; APPLICANT: Roth, Wilfred
; APPLICANT: Stenner-Liewen, Frank
; TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins
; FILE REFERENCE: P-LJ 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1383
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1380)
;
US-10-001-254-15

Query Match 100.0%; Score 294; DB 15; Length 1383;
Best Local Similarity 100.0%; Pred. No. 1.3e-74;
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1   ACATATGTGCGCTGCCTCAATGTTGGACTAAATTAGGAAGCTGTCAGAGATTTATTGATCCT 60
Db 25 ACATATGTGCGCTGCCTCAATGTTGGACTAAATTAGGAAGCTGTCAGAGATTTATTGATCCT 84

Qy 61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATGATAC 120
Db 85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATGATAC 144

Qy 121 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
Db 145 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 204

Qy 181 GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 240
Db 205 GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 264

Qy 241 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
Db 265 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 318

Column 2

RESULT 3
US-09-966-451-3
; Sequence 3, Application US/09966451
; Publication No. US20030087856A1
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF IL-1 RECEPTOR-ASSOCIATED KINASE-4 EXPRESS
; FILE REFERENCE: RTS-0324
; CURRENT APPLICATION NUMBER: US/09/966,451
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 3
; LENGTH: 2817
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (50)...(1432)
;
US-09-966-451-3

Query Match 100.0%; Score 294; DB 11; Length 2817;
Best Local Similarity 100.0%; Pred. No. 1.9e-74;
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1   ACATATGTGCGCTGCCTCAATGTTGGACTAAATTAGGAAGCTGTCAGAGATTTATTGATCCT 60
Db 74 ACATATGTGCGCTGCCTCAATGTTGGACTAAATTAGGAAGCTGTCAGAGATTTATTGATCCT 133

Qy 61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATGATAC 120
Db 134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATGATAC 193

Qy 121 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
Db 194 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 253

Qy 181 GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 240
Db 254 GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 313

Qy 241 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
Db 314 ATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 367

RESULT 4
US-10-001-254-27
; Sequence 27, Application US/10001254
; Publication No. US20030049702A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sug Hyung
; APPLICANT: Roth, Wilfred
; APPLICANT: Stenner-Liewen, Frank
; TITLE OF INVENTION: No. US20030049702A1e1 Death Domain Proteins
; FILE REFERENCE: P-LJ 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 2817
; TYPE: DNA
; ORGANISM: Homo sapien
;

```
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (50)...(1429)
US-10-001-254-27

Query Match          100.0%;  Score 294;  DB 15;  Length 2817;
Best Local Similarity 100.0%;  Pred. No. 1.9e-74;
Matches 294;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1 ACATATGTGCGCTGCCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74 ACATATGTGCGCTGCCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 133

Qy   61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGAATAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGAATAC 193

Qy  121 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  194 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 253

Qy  181 GAATTACTGTTTGACTGGGGGGCCACCACAAAATTGCACAGTTGGTGATCTCTTTTTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  254 GAATTACTGTTTGACTGGGGGGCCACCACAAAATTGCACAGTTGGTGATCTCTTTTTG 313

Qy  241 ATCCAAAAATGAATTTTTTGCTCCTGCGGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  314 ATCCAAAAATGAATTTTTTGCTCCTGCGGAGTCTTTTGCTCCCAGATGCTGTTCCC 367
```

```
RESULT 5
US-09-759-595-2
Sequence 2, Application US/09759595
Publication No. US20030059916A1
; GENERAL INFORMATION:
;   APPLICANT: Wesche, Holger
;   APPLICANT: Li, Shyun
;   APPLICANT: Tularik Inc.
;   TITLE OF INVENTION: IRAK-4: Compositions and Methods of Use
;   FILE REFERENCE: 01878I-00391OUS
;   CURRENT APPLICATION NUMBER: US/09/759,595
;   CURRENT FILING DATE: 2001-01-13
;   PRIOR APPLICATION NUMBER: US 60/176,395
;   PRIOR FILING DATE: 2000-01-13
;   NUMBER OF SEQ ID NOS: 7
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1383
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: human IL-1 receptor-associated kinase 4 (IRAK-4)
;   FEATURE:
;   OTHER INFORMATION: cDNA
;   NAME/KEY: CDS
;   LOCATION: (1)...(1383)
;   OTHER INFORMATION: human IRAK-4
US-09-759-595-2

Query Match          99.5%;  Score 292.4;  DB 11;  Length 1383;
Best Local Similarity 99.7%;  Pred. No. 3.9e-74;
Matches 293;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy    1 ACATATGTGCGCTGCCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACATATGTGCGCTGCCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 84

Qy   61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGAATAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGAATAC 144

Qy  121 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  145 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 204
```

```
Db  145 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 204

Qy  181 GAATTACTGTTTGACTGGGGGGCCACCACAAAATTGCACAGTTGGTGATCTCTTTTG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  205 GAATTACTGTTTGACTGGGGGGCCACCACAAAATTGCACAGTTGGTGATCTCTTTTG 264

Qy  241 ATCCAAAAATGAATTTTTTGCTCCTGCGGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  265 ATCCAAAAATGAATTTTTTGCTCCTGCGGAGTCTTTTGCTCCCAGATGCTGTTCCC 318
```

```
RESULT 6
US-09-833-790-149
Sequence 149, Application US/09833790
Patent No. US20020068288A1
; GENERAL INFORMATION:
;   APPLICANT: Lodes, Michael J.
;   APPLICANT: Wang, Tongtong
;   APPLICANT: Secrist, Heather
;   APPLICANT: Mohamath, Raodoh
;   APPLICANT: Indirias, Carol Y.
;   APPLICANT: Fan, Liqun
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;   AND DIAGNOSIS OF LUNG CANCER
;   FILE REFERENCE: 210121.512
;   CURRENT APPLICATION NUMBER: US/09/833,790
;   CURRENT FILING DATE: 2001-04-11
;   NUMBER OF SEQ ID NOS: 440
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149
;   LENGTH: 501
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(501)
;   OTHER INFORMATION: n = A,T,C or G
US-09-833-790-149

Query Match          82.2%;  Score 241.6;  DB 9;  Length 501;
Best Local Similarity 97.3%;  Pred. No. 1.1e-59;
Matches 288;  Conservative 0;  Mismatches 4;  Indels 4;  Gaps 4;

Qy    1 ACATATGTGCGCTGCCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   45 ACATATGTGCGCTGCCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 104

Qy   61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGAATAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  105 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGAATAC 164

Qy  121 AATC-AGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTC 179
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  165 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTC-CTTCAAACTGGAAAAAGTCCCACTTC 223

Qy  180 TGAATTACTGTTTGACT-GGGGCCACCACAAAATTGCACAGTTGGTGATCTTTTGGATCTTT 238
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db  224 TTGAATTACTGTTTGACTGGGGGGCCACCACAAAATTGCACAGTTGGTGATCTTTTGGATCTTT 283

Qy  239 TGATCCAAAAATGAATTTTTTGCTCCTGCGGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  284 TGATCCAAAAATGAA-TTTTTGCTCCTGCGGAGTCTTTTGCTCCCAGATGCTGTTCCC 338
```

```
RESULT 7
US-09-759-595-4
Sequence 4, Application US/09759595
Publication No. US20030059916A1
; GENERAL INFORMATION:
;   APPLICANT: Wesche, Holger
;   APPLICANT: Li, Shyun
;   APPLICANT: Tularik Inc.
;   TITLE OF INVENTION: IRAK-4: Compositions and Methods of Use
```

; FILE REFERENCE: 018781-00391OUS
; CURRENT APPLICATION NUMBER: US/09/759,595
; CURRENT FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 60/176,395
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 4
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: murine IL-1 receptor-associated kinase 4 (IRAK-4)
; FEATURE:
; OTHER INFORMATION: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (163)..(1542)
; OTHER INFORMATION: murine IRAK-4
US-09-759-595-4

Query Match 69.5%; Score 204.4; DB 11; Length 1542;
Best Local Similarity 81.0%;
Matches 238; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy   1 ACATATGTGCGCTGCCTCAATGTTGGACTAATGTTAGGAAGCTGTCAGAATTTTATTGATCCT 60
         |||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 187 ACATACATACGGCAACCTTAATGTGGGGAGATCCTTAGGAAGCTGTCCGGATTTATTGATCCT 246

Qy   61 CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTATTATTAAAAAAACCATCTGGTGATGATAGATAC 120
         |||||||| ||||||||| |||||| ||||||||| |||||| |||||||| | |||| |||||
Db 247 CAAGAAGGGTGGGAAGAAATTAGCAGTAGCTATCAAAAAACCGTCTCGGGACGACAGAGATAC 306

Qy 121 AATCAGTTTCACATAAGGAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
         |||||||| |||| ||||||| || |||||| |||||||| ||| |||| |||||| | |
Db 307 AATCAGTTCCATATAAGGAGGATTCGAAGCCTTACTTCAGACCCGGGAAGAGCCCACCTGT 366

Qy 181 GAATTACTGTTTGACTGGGCGCCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTG 240
         ||| | |||||||||| |||||||||||| |||||||||||| | | ||| |||| ||||||
Db 367 GAACTGCTCGTTTGACTCGGGCCACCACAGTTGCACAGTTGGGGAATCTACTG 426

Qy 241 ATCCAAAAATGAATTTTTTGCTCCTGCGAGTGCTTTTGGGAGATGGCTTGTTCCC 294
         || |||||| |||| |||||||||||| || |||||||| || ||||| ||||||
Db 427 GTCCAGAATTGAGCTGTTTTGCCCCCCGGCCACTCTCCTGCGCGGACGATGCCCGGTTCCC 480

RESULT 8
US-09-966-451-10
; Sequence 10, Application US/09966451
; Publication No. US20030087856A1
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF IL-1 RECEPTOR-ASSOCIATED KINASE-4 EXPRES
; FILE REFERENCE: RTS-0324
; CURRENT APPLICATION NUMBER: US/09/966,451
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 10
; LENGTH: 31000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-966-451-10

Query Match 51.6%; Score 151.8; DB 11; Length 31000;
Best Local Similarity 95.7%; Pred. No. 7.4e-33;
Matches 156; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 123 TCAGTTTCACATAAGGAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGA 182
         | |||| ||||| ||||||| ||||||||||||||||||||| |||||||||||| ||||
Db 12953 TTACTTACTTTTAAGGAGGCATTACTTCAAACTGGGAAAAAGTCCCACTTCTGA 13012

Qy 183 ATTACTGTTTGACTAAGGGGGGGCACCACAAAATTGCACAGTTGGTGATCTTGTGGATCTTTTGAT 242
         |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13013 ATTACTGTTTGACTGGGCGGGCGGACCACCAAAATTGCACAGTTGGTGATCTTGTGTGGATCTTTTGAT 13072

Qy 243 CCAAAATGAATTTTTTGCTCCTGCGGGCGGCCACCACAAAATTGCACAGTCTTTTGCTCCCAGAT 285
         ||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||||||||| |
Db 13073 CCAAAATGAATTTTTTTGCTCCTGCGGGCGGGACCACCAAAATTGCACAGTCTTTTGCTCCCAGGT 13115

RESULT 9
US-10-001-254-25
; Sequence 25, Application US/10001254
; Publication No. US20030497021A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Lee, Sug Hyung
; APPLICANT: Roth, Wilfried
; APPLICANT: Stenner-Liewen, Frank
; TITLE OF INVENTION: NO. US20030497021A1e1 Death Domain Proteins
; FILE REFERENCE: P-LJ 5037
; CURRENT APPLICATION NUMBER: US/10/001,254
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/301,889
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/715,893
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 211
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(177)
US-10-001-254-25

Query Match 46.8%; Score 137.6; DB 15; Length 211;
Best Local Similarity 97.2%; Pred. No. 8.5e-30;
Matches 140; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy   1 ACATATGTGCGCTGCCTCAATGTTGGACTAATGTTAGGAAGCTGTCAGAATTTTATTGATCCT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACATATGTGCGCTGCCTCAATGTTGGACTAATGTTAGGAAGCTGTCAGAATTTTATTGATCCT 84

Qy   61 CAAGAGAGGATGGGAAGAAGTTAGCTGTAGCTATTATTAAAAAAACCATCTGGTGATGATAGATAC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAAGAGAGGATGGGAAGAAGTTAGCTGTAGCTATTATTAAAAAAACCATCTGGTGATGATAGATAC 144

Qy 121 AATCAGTTTCACATAAGGAGGATTT 144
         |||||||||||||||||| ||||
Db 145 AATCAGTTTCACATAAGATGCTGT 168

RESULT 10
US-09-887-586A-51
; Sequence 51, Application US/09887586A
; Patent No. US20020094556A1
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: NO. US20020094556A11, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/887,586A
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/398,395
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22

```
;    PRIOR APPLICATION NUMBER: 60/150,262
;    PRIOR FILING DATE: 1999-08-23
;    NUMBER OF SEQ ID NOS: 58
;    SOFTWARE: FastSEQ for Windows Version 3.0
;    SEQ ID NO 51
;      LENGTH: 2024
;      TYPE: DNA
;      ORGANISM: Lycopersicon esculentum
;      FEATURE:
;      NAME/KEY: CDS
;      LOCATION: (32)...(1675)
;      OTHER INFORMATION: VFNT germacrene C synthase
  US-09-887-586A-51

Query Match        13.7%;  Score 40.4;  DB 9;  Length 2024;
Best Local Similarity   50.5%;  Pred. No. 0.31;
Matches 98;  Conservative  0;  Mismatches 96;  Indels  0;  Gaps  0;

Qy    45 AGATTTTATTGATCCTCAAGAAGGAGGAGAAGAAGTTAGCTGTAGCTATTAAAAAACCATC 104
            |||  ||  |  ||| ||||   | |||||  ||||||||| ||| ||| |||||||||||
Db   754 AGAGCTTAGTGATCTTACAAGGTGGTGGAAAGATTTGGATTTTGCAAATAAATATCCATA 813

Qy   105 TGGTGATGATAGATACAAATCAGTTTCACATAAGGAGGAGATTGAAGCATTACTTCAAACTGG 164
Db   814 TGCAAGAGACAGGTTGGTTGAGTGTTACTTCTGGATATTAGGAGTGTTATTTGAGCCAAA 873

Qy   165 AAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGGCACCACAAATTGCACAGTTGGTGA 224
Db   874 ATATAGTCGTGCGGAGAAAAATGATGACAAAAGTACTCAACCTGACCTCCATTATTGACGA 933

Qy   225 TCTTGTGTGGATCTTT 238
Db   934 CACTTTTGATGCTT 947

RESULT 11
US-09-903-012-51
;   Sequence 51, Application US/09903012
;   Patent No. US20020094557A1
;   GENERAL INFORMATION:
;      APPLICANT: Chappell, Joseph
;      APPLICANT: No. US2002009455A11, Joseph P.
;      APPLICANT: Starks, Courtney M.
;      APPLICANT: Manna, Kathleen R.
;      TITLE OF INVENTION: SYNTHASES
;      FILE REFERENCE: 07678-025001
;      CURRENT APPLICATION NUMBER: US/09/903,012
;      CURRENT FILING DATE: 2001-07-11
;      PRIOR APPLICATION NUMBER: 09/398,395
;      PRIOR FILING DATE: 1999-09-17
;      PRIOR APPLICATION NUMBER: 60/100,993
;      PRIOR FILING DATE: 1998-09-18
;      PRIOR APPLICATION NUMBER: 60/130,628
;      PRIOR FILING DATE: 1999-04-22
;      PRIOR APPLICATION NUMBER: 60/150,262
;      PRIOR FILING DATE: 1999-08-23
;      NUMBER OF SEQ ID NOS: 58
;      SOFTWARE: FastSEQ for Windows Version 3.0
;      SEQ ID NO 51
;        LENGTH: 2024
;        TYPE: DNA
;        ORGANISM: Lycopersicon esculentum
;        FEATURE:
;        NAME/KEY: CDS
;        LOCATION: (32)...(1675)
;        OTHER INFORMATION: VFNT germacrene C synthase
  US-09-903-012-51

Query Match        13.7%;  Score 40.4;  DB 9;  Length 2024;
Best Local Similarity   50.5%;  Pred. No. 0.31;
Matches 98;  Conservative  0;  Mismatches 96;  Indels  0;  Gaps  0;

Qy    45 AGATTTTATTGATCCTCAAGAAGGAGGAGAAGAAGTTAGCTGTAGCTATTAAAAAACCATC 104
Db   754 AGAGCTTAGTGATCTTACAAGGTGGTGGAAAGATTTGGATTTTGCAAATAAATATCCATA 813
Qy   105 TGGTGATGATAGATACAAATCAGTTTCACATAAGGAGGAGATTGAAGCATTACTTCAAACTGG 164
Db   814 TGCAAGAGACAGGTTGGTTGAGTGTTACTTCTGGATATTAGGAGTGTTATTTGAGCCAAA 873
Qy   165 AAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGGCACCACAAATTGCACAGTTGGTGA 224
Db   874 ATATAGTCGTGCGGAGAAAAATGATGACAAAAGTACTCAACCTGACCTCCATTATTGACGA 933
Qy   225 TCTTGTGTGGATCTTT 238
Db   934 CACTTTTGATGCTT 947
```

```
RESULT 12
US-09-900-797-51
;   Sequence 51, Application US/09900797
;   Publication No. US20030087406A1
;   GENERAL INFORMATION:
;      APPLICANT: Chappell, Joseph
;      APPLICANT: No. US20030087406A11, Joseph P.
;      APPLICANT: Starks, Courtney M.
;      APPLICANT: Manna, Kathleen R.
;      TITLE OF INVENTION: SYNTHASES
;      FILE REFERENCE: 07678-025001
;      CURRENT APPLICATION NUMBER: US/09/900,797
;      CURRENT FILING DATE: 2001-07-06
;      PRIOR APPLICATION NUMBER: 60/130,628
;      PRIOR FILING DATE: 1999-09-17
;      PRIOR APPLICATION NUMBER: 09/398,395
;      PRIOR FILING DATE: 1999-04-22
;      PRIOR APPLICATION NUMBER: 60/150,262
;      PRIOR FILING DATE: 1999-08-23
;      NUMBER OF SEQ ID NOS: 58
;      SOFTWARE: FastSEQ for Windows Version 3.0
;      SEQ ID NO 51
;        LENGTH: 2024
;        TYPE: DNA
;        ORGANISM: Lycopersicon esculentum
;        FEATURE:
;        NAME/KEY: CDS
;        LOCATION: (32)...(1675)
;        OTHER INFORMATION: VFNT germacrene C synthase
  US-09-900-797-51

Query Match        13.7%;  Score 40.4;  DB 11;  Length 2024;
Best Local Similarity   50.5%;  Pred. No. 0.31;
Matches 98;  Conservative  0;  Mismatches 96;  Indels  0;  Gaps  0;

Qy    45 AGATTTTATTGATCCTCAAGAAGGAGGAGAAGAAGTTAGCTGTAGCTATTAAAAAACCATC 104
Db   754 AGAGCTTAGTGATCTTACAAGGTGGTGGAAAGATTTGGATTTTGCAAATAAATATCCATA 813
Qy   105 TGGTGATGATAGATACAAATCAGTTTCACATAAGGAGGAGATTGAAGCATTACTTCAAACTGG 164
Db   814 TGCAAGAGACAGGTTGGTTGAGTGTTACTTCTGGATATTAGGAGTGTTATTTGAGCCAAA 873
Qy   165 AAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGGCACCACAAATTGCACAGTTGGTGA 224
Db   874 ATATAGTCGTGCGGAGAAAAATGATGACAAAAGTACTCAACCTGACCTCCATTATTGACGA 933
Qy   225 TCTTGTGTGGATCTTT 238
Db   934 CACTTTTGATGCTT 947

RESULT 13
US-09-938-842A-4518
;   Sequence 4518, Application US/0993842A
;   Patent No. US20020160378A1
;   GENERAL INFORMATION:
```

```
RESULT 13
US-09-903-012-51
Query Match        13.7%;  Score 40.4;  DB 9;  Length 2024;
Best Local Similarity   50.5%;  Pred. No. 0.31;
Matches 98;  Conservative  0;  Mismatches 96;  Indels  0;  Gaps  0;

Qy    45 AGATTTTATTGATCCTCAAGAAGGAGGAGAAGAAGTTAGCTGTAGCTATTAAAAAACCATC 104
```

```
;   APPLICANT: Harper, Jeff
;   APPLICANT: Kreps, Joel
;   APPLICANT: Wang, Xun
;   APPLICANT: Zhu, Tong
;   TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;     TITLE OF INVENTION: SAME, AND METHODS OF USE
;   FILE REFERENCE: SCRIP1300-3
;   CURRENT APPLICATION NUMBER: US/09/938,842A
;   CURRENT FILING DATE: 2001-08-24
;   PRIOR APPLICATION NUMBER: US 60/227,866
;   PRIOR FILING DATE: 2000-08-24
;   PRIOR APPLICATION NUMBER: US 60/264,647
;   PRIOR FILING DATE: 2001-01-16
;   PRIOR APPLICATION NUMBER: US 60/300,111
;   PRIOR FILING DATE: 2001-06-22
;   NUMBER OF SEQ ID NOS: 5379
;   SEQ ID NO 4518
;     LENGTH: 2000
;     TYPE: DNA
;     ORGANISM: Arabidopsis thaliana
US-09-938-842A-4518

Query Match        12.2%;    Score 36;  DB 10;  Length 2000;
Best Local Similarity   58.3%;    Pred. No. 5.8;
Matches  63; Conservative   0; Mismatches   45; Indels   0; Gaps   0;

Qy    111 TGATAGATACAAATCAGTTCACATAAGGAGAGAGATTGAGAGCATTACTTCAAACTGGAAAAAG 170
          ||||||||||||  ||||  || |||||||||||||||||||  ||||||||||||||||||
Db    1374 TTATTGAAAGTAAAGTAGTATTTTTAATTAGGAGAGATTTTAAAATTAAATGTCATTATTCTAAAATG 1433

Qy    171 TCCCACTTCTGAAATTACTCGTTTGACTGGGGGCACCACAAAATTGCACAGT 218
          ||  || ||| |||||||||||||||||||||||   ||||||||||||||
Db    1434 TACAATATATACATTATTTTTTGAAGATTTTTACAATAAAATTTCAAAGT 1481

;   RESULT 14
US-09-960-858-2
;   Sequence 2, Application US/09960858
;   Publication No. US20030138777A1
;   GENERAL INFORMATION:
;   APPLICANT: Evans, Glen
;   TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
;     TITLE OF INVENTION: USE
;   FILE REFERENCE: P-EA 4974
;   CURRENT APPLICATION NUMBER: US/09/960,858
;   CURRENT FILING DATE: 2001-09-20
;   NUMBER OF SEQ ID NOS: 19
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 2
;     LENGTH: 22684
;     TYPE: DNA
;     ORGANISM: M. genitalium
US-09-960-858-2

Query Match        11.5%;    Score 33.8;  DB 13;  Length 22684;
Best Local Similarity   53.4%;    Pred. No. 80;
Matches  71; Conservative   0; Mismatches   62; Indels   0; Gaps   0;

Qy    143 TTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCA 202
          |||||||||||| ||||||  || ||||||||||||||||||||||||| |||||| ||
Db    17554 TTGAAGCAAATTGTTGAAACCAAAAACCAAATCCTTGAATTGAATCATTATTTGACAGGATTG 17613

Qy    203 CCACAAATTGCACAGTTGGTGATCTTTGGGATCTTTTGATCCAAAATGAATTTTTTTGCTC 262
          ||||||||||||||||||  | |||| | | | |||||| | ||||  || |||||||||
Db    17614 TTAATCGCTACTCAATTACTCCTATAGTTGATCCTGAAACACAAAAAACTATTGTTAGAAG 17673

Qy    263 CTGCGAGTCTTTTT 275
          ||||||||||| |
Db    17674 CTAACAGTCTTAT 17686
```

```
;   Sequence 2, Application US/09960870
;   Publication No. US20030134281A1
;   GENERAL INFORMATION:
;   APPLICANT: Evans, Glen
;   TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
;     TITLE OF INVENTION: USE
;   FILE REFERENCE: P-EA 4738
;   CURRENT APPLICATION NUMBER: US/09/960,870
;   CURRENT FILING DATE: 2001-09-20
;   NUMBER OF SEQ ID NOS: 19
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 2
;     LENGTH: 22684
;     TYPE: DNA
;     ORGANISM: M. genitalium
US-09-960-870-2

Query Match        11.5%;    Score 33.8;  DB 13;  Length 22684;
Best Local Similarity   53.4%;    Pred. No. 80;
Matches  71; Conservative   0; Mismatches   62; Indels   0; Gaps   0;

Qy    143 TTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCA 202
          |||||||||||| ||||||  || ||||||||||||||||||||||||| |||||| ||
Db    17554 TTGAAGCAAATTGTTGAAACCAAAAACCAAATCCTTGAATTGAATCATTATTTGACAGGATTG 17613

Qy    203 CCACAAATTGCACAGTTGGTGATCTTTGGGATCTTTTGATCCAAAATGAATTTTTTTGCTC 262
          ||||||||||||||||||  | |||| | | | |||||| | ||||  || |||||||||
Db    17614 TTAATCGCTACTCAATTACTCCTATAGTTGATCCTGAAACACAAAAAACTATTGTTAGAAG 17673

Qy    263 CTGCGAGTCTTTTT 275
          ||||||||||| |
Db    17674 CTAACAGTCTTAT 17686
```

Search completed: January 18, 2004, 01:19:33
Job time : 218 secs

RESULT 15
US-09-960-870-2

OM nucleic - nucleic search, using sw model

Run on: January 17, 2004, 21:47:26 ; Search time .1514 Seconds
                                             (without alignments)
                                             7944.142 Million cell updates/sec

                                             GenCore version 5.1.6
                                             Copyright (c) 1993 - 2004 Compugen Ltd.

Title: US-10-001-254-5
Perfect score: 294
Sequence: 1 acatatgtgcgtgcctccaa.........tgctccaagatgctgtgtccc 294

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl1.*

1: gb_ba:*
2: gb_btg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_Om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_vi:*
31: em_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htg_hum:*
40: em_htgo_hum:*
41: em_htgo_mus:*
42: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Match Length	DB ID		Description
	1	294	100.0	294	6	AX431296	AX431296 Sequence
	2	294	100.0	833	6	AR223870	AR223870 Sequence
	3	294	100.0	1383	6	AX431306	AX431306 Sequence
	4	294	100.0	2817	6	AX431318	AX431318 Sequence
	5	294	100.0	2817	9	AF155118	AF155118 Homo sapi
	6	294	100.0	2820	6	AK000528	AK000528 Homo sapi
	7	292.4	99.5	1383	6	AX196260	AX196260 Sequence
	8	292.4	99.5	1383	9	AF445802	AF445802 Homo sapi
	9	292.4	99.5	1629	6	BC013316	BC013316 Homo sapi
	10	241.6	82.2	501	6	AX321132	AX321132 Sequence
	11	204.4	69.5	1542	6	AX196262	AX196262 Sequence
	12	204.4	69.5	1542	10	AF445803	AF445803 Sequence
	13	204.4	69.5	2431	10	BC051676	BC051676 Mus muscu
	14	151.8	51.6	33033	9	AY186092	AY186092 Homo sapi
c	15	151.8	51.6	118572	2	AC093012	AC093012 Homo sapi
c	16	151.8	51.6	165868	2	AC021719	AC021719 Homo sapi
c	17	151.8	51.6	309582	2	AC025567	AC025567 Homo sapi
	18	137.6	46.8	211	6	AX431316	AX431316 Sequence
	19	137.6	46.8	2213	6	BD155790	BD155790 Primer fo
	20	137.6	46.8	2213	9	AK027301	AK027301 Homo sapi
	21	110	37.4	264657	2	AC129390	AC129390 Rattus no
	22	94.8	32.2	1719	5	BC045381	BC045381 Danio rer
c	23	61.4	20.9	244696	2	AC123081	AC123081 Rattus no
c	24	61.4	20.9	252400	2	AC094859	AC094859 Rattus no
	25	54.2	18.4	756	11	BV075713	BV075713 S212P6044
c	26	46.8	15.9	88858	2	AC134788_3	AC134788_3
c	27	46.8	15.9	110000	2	AC103663_1	AC103663_1
c	28	43	14.6	9990	1	AE010646	AE010646 Fusobacte
	29	42.8	14.6	574	11	BV034947	BV034947 S212P6037
c	30	42.6	14.5	1694	9	HUMIMPA6	HUMIMPA6 Homo sapi
	31	42.6	14.5	2164	9	BC047409	BC047409 Homo sapi
c	32	42.4	14.4	2000	8	AF279456	AF279456 Lycopersi
	33	41.6	14.1	68858	9	AL139425	AL139425 Human DNA
c	34	41.6	14.1	185158	9	AC008732	AC008732 Homo sapi
	35	41	13.9	160851	5	AL954333	AL954333 Zebrafish
	36	40.8	13.9	943	11	CNS06ISR	CNS06IST
c	37	40.8	13.9	980	11	CNS06TK2	AL400707 T7 end of
	38	40.8	13.9	6095	8	SBA435550	AJ435550 Saccharom
	39	40.4	13.7	1879	8	AR183320	AR183320 Lycopersi
	40	40.4	13.7	1879	8	AF035630	AF035630 Lycopersi
	41	40.4	13.7	2003	8	AF279454	AF279454 Lycopersi
	42	40.4	13.7	2018	8	AF279453	AF279453 Lycopersi
	43	40.4	13.7	2024	8	AR183921	AR183921 Sequence
	44	40.4	13.7	2024	6	AR240709	AR240709 Sequence
	45	40.4	13.7	2024	6	AR266984	AR266984 Sequence

ALIGNMENTS

RESULT 1
AX431296
LOCUS   AX431296   294 bp   DNA   linear   PAT 28-JUN-2002
DEFINITION Sequence 5 from Patent WO0240680.
ACCESSION   AX431296
VERSION   AX431296.1 GI:21656165
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
            Roth,W. and Stenner-Liewen,F.
TITLE   Novel death domain proteins

Pred. No. is the number of results predicted by chance to have a

JOURNAL   Patent: WO 0240680-A 5 23-MAY-2002;
        BURNHAM INST (US)
FEATURES
    source   Location/Qualifiers
           1..294
           /organism="Homo sapiens"
           /mol_type="genomic DNA"
           /db_xref="taxon:9606"
    CDS   <1..294
           /note="unnamed protein product"
           /codon_start=1
           /protein_id="CAD37274.1"
           /db_xref="GI:21656166"
           /translation="TYVRCLNVGLIRKLSDFIDPQGWKKLAVAIKPSGDRYNQFH
           IRRFEALLQTGKSSPTSELLFDWGTTNCTVGDLIVDLLIQNEFFAPASLLLPDAVP"
BASE COUNT   84 a   55 c   62 g   93 t
ORIGIN

Query Match   100.0%;   Score 294;   DB 6;   Length 294;
Best Local Similarity   100.0%;   Pred. No. 1.4e-64;
Matches 294;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy      1 ACATATGTGCGCTGCCTCCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60

Qy     61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 120

Qy    121 AATCAGTTTCACATAAGGAGGATTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AATCAGTTTCACATAAGGAGGATTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 180

Qy    181 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTTG 240

Qy    241 ATCCAAAAATGAATTTTTTTGCTCCTCGCGAGTCTCTTTTGCTCCCAGATGCTGTTCCC 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATCCAAAAATGAATTTTTTTGCTCCTCGCGAGTCTCTTTTGCTCCCAGATGCTGTTCCC 294
```

RESULT 2
LOCUS   AR223870   833 bp   DNA   linear   PAT 26-SEP-2002
DEFINITION   Sequence 10 from patent US 6440663.
ACCESSION   AR223870
VERSION   AR223870.1   GI:23332452
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM   Unknown.
          Unclassified.
REFERENCE   1 (bases 1 to 833)
  AUTHORS   Scanlan,M.J., Stockert,E., Chen,Y.-T., Old,L.J., Jager,E. and
         Knuth,A.
  TITLE   Renal cancer associated antigens and uses therefor
  JOURNAL   Patent: US 6440663-A 10 27-AUG-2002;
FEATURES
    source   Location/Qualifiers
           1..833
           /organism="unknown"
BASE COUNT   273 a   155 c   179 g   226 t
ORIGIN

Query Match   100.0%;   Score 294;   DB 6;   Length 833;
Best Local Similarity   100.0%;   Pred. No. 1.3e-64;
Matches 294;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy      1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     74 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 133

Qy     61 CAAGAAGGATGGAAGAAGTTAGCTGTTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134 CAAGAAGGATGGAAGAAGTTAGCTGTTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 193

Qy    121 AATCAGTTTCACATAAGGAGGATTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194 AATCAGTTTCACATAAGGAGGATTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 253

Qy    181 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTTG 313

Qy    241 ATCCAAAAATGAATTTTTTTGCTCCTCGCGAGTCTCTTTTGCTCCCAGATGCTGTTCCC 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 ATCCAAAAATGAATTTTTTTGCTCCTCGCGAGTCTCTTTTGCTCCCAGATGCTGTTCCC 367
```

RESULT 3
LOCUS   AX431306   1383 bp   DNA   linear   PAT 28-JUN-2002
DEFINITION   Sequence 15 from Patent WO0240680.
ACCESSION   AX431306
VERSION   AX431306.1   GI:21656175
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
         Roth,W. and Stenner-Liewen,F.
  TITLE   Novel death domain proteins
  JOURNAL   Patent: WO 0240680-A 15 23-MAY-2002;
         BURNHAM INST (US)
FEATURES
    source   Location/Qualifiers
           1..1383
           /organism="Homo sapiens"
           /mol_type="genomic DNA"
           /db_xref="taxon:9606"
    CDS   1..1383
           /note="unnamed protein product"
           /codon_start=1
           /protein_id="CAD37279.1"
           /db_xref="GI:21656176"
           /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQGWKKLAVAIKPSG
           DRYNQFHIRRFEALLQTGKSSPTSELLFDWGTTNCTVGDLIVDLLIQNEFFAPASLLLP
           DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSYMPDSSSPENKSLEV
           SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEBGRGVVYKGVVNNTTVAVKKLAAM
           VDITTEELKQQFPDQEIKVMAKCQHENLVELLGFSSSDGDDLCLVYVMPNGSLLDRLSC
           LGGTPPLSWHMRCKIAQGAAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
           ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
           PQLLDIKEEIEDBEKTIEDYIDKKWNDADSTSVEAMYSVASQCLHEKKNKRPIKKV
           QQLLQEMTAS"
BASE COUNT   463 a   243 c   283 g   394 t
ORIGIN

Query Match   100.0%;   Score 294;   DB 6;   Length 1383;
Best Local Similarity   100.0%;   Pred. No. 1.2e-64;
Matches 294;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy      1 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 ACATATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 84

Qy     61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTGATGATAGATAC 144

Qy    121 AATCAGTTTCACATAAGGAGGATTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 AATCAGTTTCACATAAGGAGGATTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 204

Qy    181 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 GAATTACTGTTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCTTGTGGATCTTTTTG 264
```

RESULT 4

```
Qy    241 ATCCAAAATGAATTTTTTTGCTCCTGCCGAGTCTTTTTGCTCCCCAGATGCTGTTCCC 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 ATCCAAAATGAATTTTTTTGCTCCTGCCGAGTCTTTTTGCTCCCCAGATGCTGTTCCC 318
```

LOCUS   AX431318   2817 bp   DNA   linear   PAT 28-JUN-2002
DEFINITION Sequence 27 from Patent WO0240680.
ACCESSION   AX431318
VERSION   AX431318.1 GI:21156187
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2817)
  AUTHORS   Pawlowski,K., Fiorentino,L., Godzik,A., Lee,S.H., Reed,J.C.,
            Roth,W. and Stenner-Liewen,F.
  TITLE   Novel death domain proteins
  JOURNAL   Patent: WO 0240680-A 27 23-MAY-2002;
            BURNHAM INST (US)
FEATURES   Location/Qualifiers
     source   1..2817
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS   50..1432
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD37285.1"
                     /db_xref="GI:21156188"
                     /translation="MNKPIIPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKPSG
                     DDRYNQFHIRRFEALLQTGKSPSTSELLFDWGTTNCTVGDLVDLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGVVNNTTVAVKKLAAM
                     VDITTEELKQQFDQEIKVWAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKPAQTVWTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLLDKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSGASQCRHEKNKSPDIKKV
                     HQILLQEMTAS"
BASE COUNT   912 a   547 c   586 g   772 t
ORIGIN

Query Match   100.0%; Score 294; DB 6; Length 2817;
Best Local Similarity   100.0%;
Matches 294; Conservative   0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 ACATATGTGCGCTGCCCTCAATGTTGGACTAATTAGGAGAGCTGTCAGATTTTATTGATCCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     74 ACATATGTGCGCTGCCCTCAATGTTGGACTAATTAGGAGAGCTGTCAGATTTTATTGATCCT 133

Qy     61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTTGATGATAGATAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTTGATGATAGATAC 193

Qy    121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 253

Qy    181 GAATTACTGTTTGACTGGGGACCACCAAAATTGCACAGTTGGTGGTGATCTTGTGGATCTTTTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 GAATTACTGTTTGACTGGGGACCACCAAAATTGCACAGTTGGTGGTGATCTTGTGGATCTTTTG 313

Qy    241 ATCCAAAATGAATTTTTTTGCTCCTGCCGAGTCTTTTTGCTCCCCAGATGCTGTTCCC 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 ATCCAAAATGAATTTTTTTGCTCCTGCCGAGTCTTTTTGCTCCCCAGATGCTGTTCCC 367
```

RESULT 5

LOCUS   AF155118   2817 bp   mRNA   linear   PRI 05-JAN-2000
DEFINITION Homo sapiens putative protein kinase NY-REN-64 antigen mRNA,
            complete cds.
ACCESSION   AF155118
VERSION   AF155118.1 GI:5360130
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2817)
  AUTHORS   Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
            Jongeneel,V., Gure,A.O., Jager,D., Jager,E., Knuth,A., Chen,Y.-T.
            and Old,L.J.
  TITLE   Antigens recognized by autologous antibody in patients with
            renal-cell carcinoma
  JOURNAL   Int. J. Cancer 83 (4), 456-464 (1999)
   MEDLINE   99438124
    PUBMED   10508479
REFERENCE   2 (bases 1 to 2817)
  AUTHORS   Scanlan,M.J., Gordan,J.D., Williamson,B., Stockert,E., Bander,N.H.,
            Jongeneel,V., Gure,A.O., Jager,D., Jager,E., Knuth,A., Chen,Y.-T.
            and Old,L.J.
  TITLE   Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Ludwig Institute, Sloan-Kettering
            Institute, 1275 York Ave., New York, NY 10021, USA
FEATURES   Location/Qualifiers
     source   1..2817
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="1973/10-4"
                     /cell_type="renal cell carcinoma"
     CDS   50..1432
                     /codon_start=1
                     /product="putative protein kinase NY-REN-64 antigen"
                     /protein_id="AAD42884.1"
                     /db_xref="GI:5360131"
                     /translation="MNKPIIPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKPSG
                     DDRYNQFHIRRFEALLQTGKSPSTSELLFDWGTTNCTVGDLVDLIQNEFFAPASLLLP
                     DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGVVNNTTVAVKKLAAM
                     VDITTEELKQQFDQEIKVWAKCQHENLVELLGFSSDGDDLCLVYVYMPNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKPAQTVWTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLLDKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSGASQCRHEKNKSPDIKKV
                     HQILLQEMTAS"
BASE COUNT   912 a   547 c   586 g   772 t
ORIGIN

Query Match   100.0%; Score 294; DB 9; Length 2817;
Best Local Similarity   100.0%;
Matches 294; Conservative   0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 ACATATGTGCGCTGCCCTCAATGTTGGACTAATTAGGAGAGCTGTCAGATTTTATTGATCCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     74 ACATATGTGCGCTGCCCTCAATGTTGGACTAATTAGGAGAGCTGTCAGATTTTATTGATCCT 133

Qy     61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTTGATGATAGATAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAAACCATCTGGTTGATGATAGATAC 193

Qy    121 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194 AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCT 253

Qy    181 GAATTACTGTTTGACTGGGGACCACCAAAATTGCACAGTTGGTGGTGATCTTGTGGATCTTTTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 GAATTACTGTTTGACTGGGGACCACCAAAATTGCACAGTTGGTGGTGATCTTGTGGATCTTTTG 313

Qy    241 ATCCAAAATGAATTTTTTTGCTCCTGCCGAGTCTTTTTGCTCCCCAGATGCTGTTCCC 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 ATCCAAAATGAATTTTTTTGCTCCTGCCGAGTCTTTTTGCTCCCCAGATGCTGTTCCC 367
```

DNA of SEQ ID #S

RESULT 6
LOCUS   AK000528   2820 bp   mRNA   linear   PRI 22-FEB-2000
DEFINITION   Homo sapiens cDNA FLJ20521 fis, clone KAT10395.
ACCESSION   AK000528
VERSION   AK000528.1   GI:7020683
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
   ORGANISM   Homo sapiens
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
   AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
     Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
     Nakamura,Y., Isogai,T. and Sugano,S.
   TITLE   NEDO human cDNA sequencing project
   JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2820)
   AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
     Shibahara,T., Suzuki,Y., Tanaka,T. and Nakamura,Y.
   TITLE   Direct Submission
   JOURNAL   Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,
     University of Tokyo, Deptment of Virology; Shirokane-dai, 4-6-1,
     Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp,
     Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
     International Trade and Industry of Japan; cDNA full insert
     sequencing: Research Association for Biotechnology; cDNA library
     construction, 5'- & 3'-end one pass sequencing: Depertment of
     Virology and Human Genome Center, Institute of Medical Science,
     University of Tokyo (partly supported by Science and Technology
     Agency).
FEATURES   Location/Qualifiers
   source   1..2820
       /organism="Homo sapiens"
       /mol_type="mRNA"
       /db_xref="taxon:9606"
       /clone="KAT10395"
       /cell_line="KATO III"
       /cell_type="signet-ring cell carcinoma"
       /clone_lib="KAT"
       /note="cloning vector pME18FL3"
   CDS   30..1412
       /note="unnamed protein product"
       /codon_start=1
       /protein_id="BAA91232.1"
       /db_xref="GI:7020684"
       /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
       DDRYNQFHIRREALLGTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
       DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSYMPPDSSSPENKSLEV
       SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGVVYKGVVNNTVAVKKLAAM
       LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
       ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
       POLLLDIKEEIEDEEKTIEDYIDKKMNDADSISVEAMYSVASQCIHEKKNKRPDIKKV
       QQLLQEMTAS"
BASE COUNT   940 a   534 c   579 g   767 t
ORIGIN

Query Match   100.0%;   Score 294;   DB 9;   Length 2820;
Best Local Similarity   100.0%;   Pred. No. 1.1e-64;
Matches   294;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 ACAThATGTCGCGCTCGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 ACAThATGTCGCGCTCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 113
Qy   61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   114 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 173
Qy   121 AATCAGTTTCACATAAGGAGATTGAAGCATTACTTCAAACTGAAAAAGTCCCACTTCT 180

RESULT 7
LOCUS   AX196260   1383 bp   DNA   linear   PAT 28-AUG-2001
DEFINITION   Sequence 2 from Patent WO0151641.
ACCESSION   AX196260
VERSION   AX196260.1   GI:15386462
KEYWORDS
SOURCE   Homo sapiens (human)
   ORGANISM   Homo sapiens
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
   AUTHORS   Wesche,H. and Li,S.\
   TITLE   Irak-4: compositions and methods of use
   JOURNAL   Patent: WO 0151641-A 2 19-JUL-2001;
     Tularik Inc. (US)
FEATURES   Location/Qualifiers
   source   1..1383
       /organism="Homo sapiens"
       /mol_type="genomic DNA"
       /db_xref="taxon:9606"
       /note="human IL-1 receptor-associated kinase 4 (IRAK-4)
   CDS   1..1383
       /note="human IRAK-4"
       /codon_start=1
       /protein_id="CAC60090.1"
       /db_xref="GI:15386463"
       /translation="MNKPITPSTYVRCLNVGLIRKLSDFIDPQEGWKKLAVAIKKPSG
       DDRYNQFHIRREALLGTGKSPTSELLFDWGTTNCTVGDLVDLLIQNEFFAPASLLLP
       DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSYMPPDSSSPENKSLEV
       SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGVVYKGVVNNTVAVKKLAAM
       LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
       ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
       POLLLDIKEEIEDEEKTIEDYIDKKMNDADSISVEAMYSVASQCIHEKKNKRPDIKKV
       QQLLQEMTAS"
BASE COUNT   463 a   244 c   283 g   393 t
ORIGIN

Query Match   99.5%;   Score 292.4;   DB 6;   Length 1383;
Best Local Similarity   99.7%;   Pred. No. 3.1e-64;
Matches   293;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   1 ACAThATGTCGCGCTCGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 ACAThATGTCGCGCTCGCCTCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 84
Qy   61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   85 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 144
Qy   121 AATCAGTTTCACATAAGGAGATTGAAGCATTACTTCAAACTGAAAAAGTCCCACTTCT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   145 AATCAGTTTCACATAAGGAGATTGAAGCATTACTTCAAACTGAAAAAGTCCCACTTCT 204
Qy   181 GAATTAACTGTTTGACTGGGGCCACCACAAATTGCACAGCTGGTGACTTGTTGTGGATCTTTG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   205 GAATTAACTGTTTGACTGGGGCCACCACAAATTGCACAGCTGGTGACTTGTTGTGGATCTTTG 264
Qy   241 ATCCAAAAATGAATTTTTTGTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATCCAAAAATGAATTTTTTGTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 294

RESULT 8

```
Db    265  ATCCAAAATGAATTTTTTTGCTCCTGCCAGTCTTTTGCTCCCCAGATGCTGTGTTCCC 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy         ATCCAAAATGAATTTTTTTGCTCCTGCCAGTCTTTTGCTCCCCAGATGCTGTGTTCCC
```

Db	AF445802
Qy	
Db	Homo sapiens interleukin-1 receptor associated kinase 4 (IRAK4)
Db	mRNA, complete cds.

```
LOCUS       AF445802    1383 bp    mRNA    linear    PRI 20-APR-2002
DEFINITION  Homo sapiens interleukin-1 receptor associated kinase 4 (IRAK4)
            mRNA, complete cds.
ACCESSION   AF445802
VERSION     AF445802.1  GI:20219009
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1383)
AUTHORS     Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
TITLE       IRAK-4: a novel member of the IRAK family with the properties of an
            IRAK-kinase
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (8), 5567-5572 (2002)
MEDLINE     21957277
PUBMED      11960013
REFERENCE   2  (bases 1 to 1383)
AUTHORS     Suzuki,N., Suzuki,S., Duncan,G.S., Millar,D.G., Wada,T.,
            Mirtsos,C., Takada,H., Wakeham,A., Itie,A., Li,S., Penninger,J.M.,
            Wesche,H., Ohashi,P.S., Mak,T.W. and Yeh,W.C.
TITLE       Severe impairment of interleukin-1 and Toll-like receptor
            signaling in mice lacking IRAK-4
JOURNAL     Nature 416 (6882), 750-756 (2002)
MEDLINE     21959395
PUBMED      11923871
REFERENCE   3  (bases 1 to 1383)
AUTHORS     Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
TITLE       Direct Submission
JOURNAL     Submitted (08-NOV-2001) Biology I, Tularik Inc., 2 Corporate Drive,
            South San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1..1383
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..1383
                     /gene="IRAK4"
     CDS             1..1383
                     /gene="IRAK4"
                     /note="protein kinase"
                     /codon_start=1
                     /product="interleukin-1 receptor associated kinase 4"
                     /protein_id="AAM15772.1"
                     /db_xref="GI:20219010"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTAGDLVDLLIQNEFFAPASLLP
                     DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTVAVKKLAAM
                     VDITEELKQFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVVVMYMNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV
                     QQLLQEMTAS"
```

```
BASE COUNT    463 a    244 c    283 g    393 t
ORIGIN
```

RESULT 9

```
Db    265  ATCCAAAATGAATTTTTTTGCTCCTGCCAGTCTTTTGCTCCCCAGATGCTGTGTTCCC 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    318  ATCCAAAATGAATTTTTTTGCTCCTGCCAGTCTTTTGCTCCCCAGATGCTGTGTTCCC
```

```
LOCUS       BC013316    1629 bp    mRNA    linear    PRI 04-SEP-2001
DEFINITION  Homo sapiens, clone MGC:13330 IMAGE:4287014, mRNA, complete cds.
ACCESSION   BC013316
VERSION     BC013316.1  GI:15426431
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1629)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (31-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 19 Row: n Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7705840.
FEATURES             Location/Qualifiers
     source          1..1629
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:13330 IMAGE:4287014"
                     /tissue_type="Brain, primitive neuroectodermal"
                     /clone_lib="NIH_MGC_56"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             71..1453
                     /codon_start=1
                     /product="Unknown (protein for MGC:13330)"
                     /protein_id="AAH13316.1"
                     /db_xref="GI:15426432"
                     /translation="MNKPITPSTYVRCLNVGLIRKLSDFDPQEGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPTSELLFDWGTTNCTAGDLVDLLIQNEFFAPASLLP
                     DAVPKTANTLPSKEAITVQQKQMPFCDKDRTLMTPVQNLEQSYMPPDSSSPENKSLEV
                     SDTRFHSFSFYELKNVTNNFDERPISVGGNKMGEGGFGVVYKGYVNNTVAVKKLAAM
                     VDITEELKQFDQEIKVMAKCQHENLVELLGFSSDGDDLCLVVVMYMNGSLLDRLSC
                     LDGTPPLSWHMRCKIAQGAANGINFLHENHHIHRDIKSANILLDEAFTAKISDFGLAR
                     ASEKFAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLEIITGLPAVDEHRE
                     PQLLLDIKEEIEDEEKTIEDYIDKKMNDADSTSVEAMYSVASQCLHEKKNKRPDIKKV"
```

Middle column (Query/Db alignment):

```
Qy
Db
Qy    121  AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGAGAAAGTCCGAAAAG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145  AATCAGTTTCACATAAGGAGATTTGAAGCATTACTTCAAACTGAGAAAGTCCGAAAAG 204

Qy    181  GAATTACTGTTTGACTGGGGCACCACAAATGCACAGTTGTTGTGTGTGATCTTTTG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205  GAATTACTGTTTGACTGGGGCACCACAAATGCACAGTTGTTGTGTGATCTTTTG 264

Qy    241  ATCCAAAATGAATTTTTTTGCTCCTGCCAGTCTTTTGCTCCCCAGATGCTGTGTTCCC 294
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265  ATCCAAAATGAATTTTTTTGCTCCTGCCAGTCTTTTGCTCCCCAGATGCTGTGTTCCC 318
```

Right column:

```
Db    85   CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 144
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy         CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC

Qy    1    ACATAATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAAGCTGTCAGATTTTTATTCAGT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    25   ACATAATGTGCGCTGCCTCAATGTTGGACTAATTAGGAAAGCTGTCAGATTTTTATTCAGT 84

Qy    61   CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         CAAGAAGGATGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC
```

Query Match statistics:

```
Query Match       99.5%;    Score 292.4;  DB 9;   Length 1383;
Best Local Similarity   99.7%;    Pred. No. 3.1e-64;
Matches  293;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;
```

BASE COUNT   542 a   291 c   328 g   468 t
ORIGIN

Query Match   "QOLLQEMTAS"
Best Local Similarity   99.5%;   Conservative   99.7%;   Pred. No.3.1e-64;
Matches 293;   0;   Mismatches   1;   Indels   0;   Gaps   0;
                           Score 292.4;   DB 9;   Length 1629;

```
Qy        1 ACATATGTGCGCTGCCTCCCTCAAATGTTGGACTAATTAGGAAGCCTGTCGTCAGATTTTATTGATCCT 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       95 ACATATGTGCGCTGCCTCCCTCAAATGTTGGACTAATTAGGAAGCCTGTCGTCAGATTTTATTGATCCT 154

Qy       61 CAAGAAGGAGGAGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      155 CAAGAAGGAGATGGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 214

Qy      121 AATCAGTTTCACATAAGGAGGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      215 AATCAGTTTCACATAAGGAGGATTTACTTCAAACTGGAAAAAAAGTCCCACTTCT 274

Qy      181 GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGGTGATCTTGTGGATCTTTTG 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      275 GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGGTGATCTTGTGGATCTTTTG 334

Qy      241 ATCCAAAAATGAAATTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      335 ATCCAAAAATGAATTTTTTGCTCCTGCCAAGTCTTTTGCTCCCAGATGCTGTTCCC 388
```

RESULT 10
AX321132

LOCUS   AX321132   501 bp   DNA   linear   PAT 15-DEC-2001
DEFINITION Sequence 149 from Patent WO0171168.
ACCESSION   AX321132
VERSION   AX321132.1   GI:17904576
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Iodes,M.J., Wang,T., Mohamath,R. and Indrias,C.Y.
  TITLE   Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0171168-A 149 18-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES   Location/Qualifiers
     source   1..501
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"

BASE COUNT   166 a   102 c   97 g   134 t
ORIGIN

Query Match   "QOLLQEMTAS"
Best Local Similarity   82.2%;   Conservative   97.3%;   Pred. No.3.2e-51;
Matches 288;   0;   Mismatches   4;   Indels   4;   Gaps   4;
                           Score 241.6;   DB 6;   Length 501;

```
Qy        1 ACATATGTGCGCTGCCTCCCTCAAATGTTGGACTAATTAGGAAGCTGTCGTCAGATTTTATTGATCCT 60
            ||||||||||||||||||| ||||||||| ||||| ||||||||||| ||||||||||||||||||
Db       45 ACATATGTGCGCTGCCTCAAATGTTGGACTAATGTTGGACTAATTAGGAGACTGTCAGATTTTATTGATCCT 104

Qy       61 CAAGAAGGATGGAGAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
            ||||||| ||| ||| ||||||||||||||| |||||||| |||||||||||||||||||
Db      105 CAAGAAGGATGGGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 164

Qy      121 AATC-AGTTTCACATAAGGAGGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTC 179
            |||| ||||||||||||||||||| |||||||||||||||||||||||| |||||||||||
Db      165 AATCAAGTTTCACATAAGGAGATTTGAAGCATT-CTTCAAACTGGAAAAGCCATTC 223

Qy      180 TGAATTACTGTTTGACT-GGGGCACCACAAAATTGCACAGTTGGGTGATCTTGTGGATCTTT 238
            ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      224 TTGAATTACTGTTTGACTGGGGCACCACAAAATTGGACAGTTGGGTGATCTTGTGGATCTTT 283
```

BASE COUNT   421 a   392 c   423 g   306 t
ORIGIN

```
Qy      239 TGATCCAAAAATGAATTTTTTGCTCCTGCCGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
            |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      284 TGATCCAAAAATGAA-TTTTTGCCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 338
```

RESULT 11
AX196262

LOCUS   AX196262   1542 bp   DNA   linear   PAT 28-AUG-2001
DEFINITION Sequence 4 from Patent WO0151641.
ACCESSION   AX196262
VERSION   AX196262.1   GI:15386464
KEYWORDS   .
SOURCE   Mus sp.
  ORGANISM Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Wesche,H. and Li,S.
  TITLE   Irak-4: compositions and methods of use
  JOURNAL   Patent: WO 0151641-A 4 19-JUL-2001;
            Tularik Inc. (US)
FEATURES   Location/Qualifiers
     source   1..1542
                     /organism="Mus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10095"
                     /note="murine IL-1 receptor-associated kinase 4 (IRAK-4)
                     cDNA"
     CDS   163..1542
                     /note="murine IRAK-4"
                     /codon_start=1
                     /protein_id="CAC60091.1"
                     /db_xref="GI:15386465"
                     /translation="MNKPLTPSTYIRNLNVGILRKLSDFIDPQEGWKLAVAIKKPSG
                     DDRYNQPHIRREFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLLVQIELFRPATLLP
                     DAVPQTVKSLPREAATVAQTHGPCQEKDRTSVMPMPKLEHSCEPPDSSSPDNRSVES
                     SDTRFHSFSFHELKSITNNFDEQPASAGGNRMGEGGFGVVYKGCVNNTIVAVKKLGAM
                     VEISTEELKQQFDQEIKVMATCQHENLVELLGFSSDSDNLCLVYAYMPNGSLLDRLSC
                     LDGTPPLSWHTRCKVAQGTANGIRFLHENHHIHRDIKSANILLKDTFAKISDFGLAR
                     ASARLAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLELITGLAAVDENRE
                     PQLLLDIKEEIEDEEKTIEDYDEKMSDADPASVEAMYSAASQCLHEKKNRPDIAKV
                     QQLLQEMSA"
```

BASE COUNT   421 a   392 c   423 g   306 t
ORIGIN

Query Match   "QOLLQEMSA"
Best Local Similarity   69.5%;   Conservative   81.0%;   Pred. No.8.6e-42;
Matches 238;   0;   Mismatches   56;   Indels   0;   Gaps   0;
                           Score 204.4;   DB 6;   Length 1542;

```
Qy        1 ACATATGTGCGCTGCCTCCCTCAAATGTTGGACTAATTAGGAAGCCTGTCGTCAGATTTTATTGATCCT 60
            |||||| ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      187 ACATACATACGCAACCTTAATGTGGGGACCCTTAGGAACCTTGTGGGATTTTATTGATCCT 246

Qy       61 CAAGAAGGATGGAGAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGATGATAGATAC 120
            |||||| |||  ||||||||||||||||| ||||||||||||||||||||||||||||||
Db      247 CAAGAAGGGGTGGAAGAAATTAGCAGTAGCTTATCAAAAAAGCCGTCCCGGCGACGAGATAC 306

Qy      121 AATCAGTTTCACATAAGGAGGATTTGAAGCATTACTTCAAACTGGAAAAAAGTCCCACTTCT 180
            |||||||||| |||||||||||||| |||  |||||||||||||||||||||||||| ||
Db      307 AATCAGTTTCCATAAGGAGAGATTCGAAGAGCCTTACTTCAGACCGGAAGAGCCCACTTCT 366

Qy      181 GAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGGTGATCTTGTGGATCTTTG 240
            ||||||||| ||| |||||||||||||||||| |||||||||| ||||||||||||||||
Db      367 GAACTGCTGTTTGACTGGGGCACCACGAACTGCACAGTTGGGGACCTTGTGGATCTTCTG 426

Qy      241 ATCCAAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGCTGTTCCC 294
            ||||||||| || || ||||  ||||||||||||||||||||||||||||||||
Db      427 GTCCAGATTGAGCTGTTTCGCCCCGCCACTCTCCTGCCGGATGCCGTTCCC 480
```

RESULT 12
AF445803

```
LOCUS       AF445803    1542 bp    mRNA    linear   ROD 20-APR-2002
DEFINITION  Mus musculus interleukin-1 receptor associated kinase 4 (Irak4)
            mRNA, complete cds.
ACCESSION   AF445803
VERSION     AF445803.1  GI:20219011
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1542)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE     IRAK-4: a novel member of the IRAK family with the properties of an
            IRAK-kinase
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (8), 5567-5572 (2002)
  MEDLINE   21957277
   PUBMED   11960013
REFERENCE   2  (bases 1 to 1542)
  AUTHORS   Li,S., Strelow,A., Fontana,E.J. and Wesche,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) Biology I, Tularik Inc., 2 Corporate Drive,
            South San Francisco, CA 94080, USA
FEATURES             Location/Qualifiers
     source          1..1542
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
     gene            1..1542
                     /gene="Irak4"
     CDS             163..1542
                     /gene="Irak4"
                     /note="protein kinase"
                     /codon_start=1
                     /product="interleukin-1 receptor associated kinase 4"
                     /protein_id="AAM15773.1"
                     /db_xref="GI:20219012"
                     /translation="MNKPLTPSTYIRNLNVGILRKLSDFIDPQGWKKLAVAIKKPSG
                     DDRYNQFHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLIVQIELFAPATLLLP
                     DAVPQTVKSLPPREAATVAQTHGPCQEKDRTSVNPMKLEHSCEPPDSSSPDRSVES
                     SDRRHSFSFHELKSITNNFDEQPASAGNRMGEGGFGVVYKGCVNNTIVAVKKLGAM
                     VEISTEELKQQFDQEIKWMATCQHENLVELLGFSSDSDNLCLVYAYMPNGSLLDRLSC
                     LDGTPPLSWHTRCKVAQGTPANGIRFLHENHIIHRDIKSANILLDKDFTAKISDFGLAR
                     ASARLAQTYMTRSIVGTPAYMAPEALRGEITPKSDIYSGVVLLELITGLAAVDENR
                     PQLLLDIKEEIEDEEKTIEDYTDEKMSDADPASVEAMYSAASQCLHEKQNRRPDIAKV
                     QOLLQEMSA"
BASE COUNT      421 a    392 c    423 g    306 t
ORIGIN
```

```
Query Match       69.5%;    Score 204.4;    DB 10;    Length 1542;
Best Local Similarity   81.0%;    Pred. No. 8.6e-42;
Matches  238;  Conservative  0;  Mismatches  56;  Indels  0;  Gaps  0;

Qy        1 ACATATGTGCGCGCTGCCTCAATGTTGGACTAATTAGGAGAAGCTGTCAGATTTTATTGATTCCT 60
            ||||| ||||||||||||||||||||||||||| ||||||||||||||| | ||||||||||||
Db      187 ACATACATACGGCAACCTTAATGTGGGGGATCCTTAGGAAGCTGTCGGATTTTATTGATTCCT 246

Qy       61 CAAGAAGGGATGGAAGAAAGTTAGCTGTTAGCTGTTATTAAAAAACCATCTGGTGATGATAGAATAC 120
            ||||||||| |||||||| |||||||||||| ||||||||||||  ||||| ||||||||||||||
Db      247 CAAGAAGGGTGGAAGAAAATTAGCAGTAGCTATCAAAAAAGCCGTCCGCGACGACAGATAC 306

Qy      121 AATCAGTTCACATAAGGAGATTTGAAGCAATTACTTCAAACTGGAAAAAGTCCCACTTCT 180
            |||||||| |||||  ||||||| ||||| ||||||||| || |||| |||||||| |
Db      307 AATCAGTTCCATATAAGGAGAATTCGAAGCCTTACTTCAGACCGGGAAGAGACCCCACCTGT 366

Qy      181 GAATTACTGTTTGACTGGGGGCACCACAAATTGCACAGTTGGTGATCTTGTCGATCTTTTG 240
            |||| ||||||||||||||||| |||||  ||| ||| |||| || | |||||| ||
Db      367 GAACTGCTGTTTGACTGGGGGCCACCAGTTGGCGACCTTGTGGGATCTACTG 426

Qy      241 ATCCAAAAATGAATTTTTTTGTCTCCCTGGGGGGGTCTTTTGGGG 294
            |||| |||| |||||||||| ||| | |||||||| |||||||||
Db      427 GTCCAGATTGAGCTGTTTTGGCTCCCGCCCACTCTCCTCGCGGATGCCGTTCCC 480
```

```
RESULT 13
BC051676
LOCUS       BC051676    2431 bp    mRNA    linear   ROD 10-JUN-2003
DEFINITION  Mus musculus interleukin-1 receptor-associated kinase 4, transcript
            variant 1, mRNA (cDNA clone MGC:60994 IMAGE:30017484), complete
            cds.
ACCESSION   BC051676
VERSION     BC051676.1  GI:30545427
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2431)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE  22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2431)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Minoru Ko
            cDNA Library Preparation: Yulan Piao and Minoru Ko
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 111 Row: i Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 23943897.
FEATURES             Location/Qualifiers
     source          1..2431
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="MGC:60994 IMAGE:30017484"
                     /tissue_type="Embryo, whole, mouse, 7.5-dpc"
```

```
COMMENT      To cite this work please use: SeattleSNPs, NHLBI HL66682 Program
             for Genomic Applications, UW-FHCRC, Seattle, WA (URL:
             http://pga.gs.washington.edu).
JOURNAL      Submitted (22-NOV-2002) Genome Sciences, University of Washington,
             1705 NE Pacific, Seattle, WA 98195, USA
TITLE        Direct Submission
AUTHORS      Rieder,M.J., Arnel,T.Z., Carrington,D.P., Ozuna,M., Kuldanek,S.A.,
             Rajkumar,N., Toth,E.J., Yi,Q. and Nickerson,D.A.
REFERENCE    1  (bases 1 to 33033)
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM     Homo sapiens
SOURCE       Homo sapiens (human)
KEYWORDS     .
VERSION      AY186092.1  GI:26000792
ACCESSION    AY186092
             gene, complete cds.
DEFINITION   Homo sapiens interleukin-1 receptor-associated kinase 4 (IRAK4)
LOCUS        AY186092               33033 bp    DNA     linear   PRI 02-DEC-2002
```

```
RESULT 14

Db    432 GTCCAGATTGAGCTGTTTGCCCCGCCCACTCTCCTGCCGGATGCCGTTCCC 485
          |||||||||| |||||||||| ||| |||| |||||||| ||||||
Qy    241 ATCCAAAAATGAATTTTTTGCTCCTCGCGAGTGTCTTTTGTCGTTCCC 294
          |||  | || | || |||| |||| |||| ||| ||| || |||
Db    372 GAACTGCTGTTTGACTGGGGCCACCACCAGTTGGCCGACCTTGTGGACTCACTG 431
          ||||| |||||||  ||| |||| |||
Qy    181 GAATTACTGTTTGACTGGGGGCGACCACCAAAATTGCACACAGTTGGTGATCTTGTGGGATCTTTTG 240
          || |||| || |||| || |||||||||||  |||  ||| ||||||||||
Db    312 AATCAGTTCCATATAAAGGAGAATTCGAAGCCTTACTTCAGACGCGGGAAGAGACCCCCACCTGT 371
          ||||||| || |  || |||||||||||||||| |||| ||||
Qy    121 AATCAGTTTCACATAAGGAGGAGATTTGAAGCATTACTTCAAAAAAGGCCGTCCGGACGACAGATAC 180
          |||||| ||| |||| ||||| |||| |||||| |||||| ||| |||||
Db    252 CAAGAAGGGTGGAAGAGAAATTAGCAGTAGCTTATCAAAAAGCCATTGGGTGAGATATAGAGATAC 311
          ||||| ||  ||||| |||||| |||||||||||||||||||||||||||
Qy     61 CAAGAAGGATGGAAGAAGTTAGCTGTAGCTGTTATTAAAAAACCATCTGGTGTGATAGAGATAC 120
          ||||| |||| |||| || |||| || |||||||||||| ||||||||| ||||||||||||
Db    192 ACATACAATACGCAACCTTAATGTCGGGGATCCTTAGGAAGCTGTCGGATTTATTGATCCT 251
          |||| |||| ||| ||||  ||| |||||| ||| |||| || || ||| ||| |||
Qy      1 ACATATGTGCGGCTGCCTCAATGTTGGACTAATTAGGAAGGTCAGATTTTTATTGATCCT 60

Matches 238; Conservative  0; Mismatches 56; Indels  0; Gaps  0;
Best Local Similarity  81.0%;  Pred. No. 8.1e-42;
Query Match  69.5%;  Score 204.4;  DB 10;  Length 2431;
```

```
ORIGIN
BASE COUNT      656 a    580 c    618 g    577 t
```

FEATURES   Location/Qualifiers

```
variation       3038
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"
repeat_region   3025..3915
                /rpt_family="L2"
                /rpt_type=dispersed
variation       3025..3915
                /gene="IRAK4"
                /frequency="0.09"
                /replace="t"
variation       3023
                /gene="IRAK4"
                /replace="t"
repeat_region   2506..2807
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       2940
                /gene="IRAK4"
                /frequency="0.04"
                /replace="t"
variation       2456
                /gene="IRAK4"
                /replace="c"
variation       2365
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"
variation       2315
                /gene="IRAK4"
                /frequency="0.01"
                /replace="t"
mRNA            join(1894..1933,11040..11209,14157,14302,15117,15299,
                15849..16009,16902..16966,20567..20681,21130..21239,
                25244..25427,26599..26661,29336..29494,29617..31033)
                /gene="IRAK4"
                /product="interleukin-1 receptor-associated kinase 4"
gene            1894..31033
                /gene="IRAK4"
variation       1533..1534
                /gene="IRAK4"
                /frequency="0.93"
                /replace=""
variation       1894..31033
                /frequency="0.01"
                /replace=""
variation       1265
                /frequency="0.01"
                /replace="a"
variation       854
                /frequency="0.02"
                /replace="a"
repeat_region   708..941
                /rpt_family="L2"
                /rpt_type=dispersed
variation       642
                /frequency="0.02"
                /replace="c"
repeat_region   384..599
                /rpt_family="L2"
                /rpt_type=dispersed
variation       133
                /frequency="0.06"
                /replace="a"
variation       92..94
                /frequency="0.02"
                /replace="c"
repeat_region   89..383
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   34..88
                /rpt_family="L2"
                /rpt_type=dispersed
source          1..33033
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
```

```
gene            1..2431
                /gene="Irak4"
                /note="Vector: pSPORT1"
                /note="Synonyms: NY-REN-64, 9330209D03Rik"
                /db_xref="LocusID:266632, 9330209D03Rik"
                /db_xref="MGI:2182474"
                /clone_lib="NIA Mouse 7.5-dpc Whole Embryo cDNA Library
                (Long)"
                /lab_host="DH10B"
CDS             168..1547
                /codon_start=1
                /product="interleukin-1 receptor-associated kinase 4,
                isoform a"
                /protein_id="AAH51676.1"
                /db_xref="GI:30354528"
                /db_xref="LocusID:266632"
                /translation="MNKPLTPSTYIRNLNVGILRKLSDFIDPQEGWKKLAVAIKKPSG
                DDRYNQFHIRRFEALLQTGKSPTCELLFDWGTTNCTVGDLVDLLVQIELFAPATLLP
                DAVPQTVKSLPREAATVAQGTGPCQEKDRTSVMPMKLKEHSCEPDSSSSDNRSVES
                SDTRFHSFSFHELKSIINNFDEQPASAGGNRMGGEGVVYKGCVNNTIVAVKKLGAM
                VEISTEELKQQFDQEIKVMATCQHENLVELLGFSSDSDSDNLCLVYAYMPNGSLLDRLSC
                LDGTPPLSWHTRCKVAQGTANGIRFLHENHHIHRDIKSANILLDRDFTAKISDFGLAR
                ASARLAQTVMTSRIVGTTAYMAPEALRGEITPKSDIYSFGVVLLELITGLAAVDENRE
                PQLLLDIKEEIEDEEKTIEDYIDEKMSDADPASVEAMYSAASQCLHEKKNRRPDIAKV
                QLLQEMSA"
```

```
variation       3115..3129
                /gene="IRAK4"
                /frequency="0.72"
                /replace=""
variation       3189
                /gene="IRAK4"
                /frequency="0.01"
                /replace="g"
variation       3536
                /gene="IRAK4"
                /frequency="0.14"
                /replace="c"
variation       3573
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"
variation       3819
                /gene="IRAK4"
                /frequency="0.01"
                /replace="t"
variation       3877
                /gene="IRAK4"
                /frequency="0.17"
                /replace="t"
variation       4144
                /gene="IRAK4"
                /frequency="0.01"
                /replace="g"
variation       4438..4610
                /gene="IRAK4"
                /frequency="0.04"
                /replace="a"
repeat_region   4438..4610
                /rpt_family="MIR"
                /rpt_type=dispersed
variation       4602
                /gene="IRAK4"
                /frequency="0.01"
                /replace="g"
variation       4797
                /gene="IRAK4"
                /frequency="0.01"
                /replace="t"
variation       4972
                /gene="IRAK4"
                /frequency="0.02"
                /replace="a"
repeat_region   5759..5797
                /rpt_family="L2"
                /rpt_type=dispersed
variation       5880
                /gene="IRAK4"
                /frequency="0.02"
                /replace="g"
variation       5954
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"
variation       5964
                /gene="IRAK4"
                /frequency="0.31"
                /replace="a"
repeat_region   5964..6036
                /rpt_family="MER1_type"
                /rpt_type=dispersed
repeat_region   6023..6590
                /rpt_family="MER1_type"
                /rpt_type=dispersed
variation       6204
                /gene="IRAK4"
                /frequency="0.04"
                /replace="t"
variation       6661..6662
                /gene="IRAK4"
                /frequency="0.04"
                /replace=""
variation       6998
                /gene="IRAK4"
                /frequency="0.04"
                /replace=""

repeat_region   7015..7073
                /gene="IRAK4"
                /rpt_family="Mariner"
                /rpt_type=dispersed
repeat_region   7074..7366
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       7112
                /gene="IRAK4"
                /frequency="0.01"
                /replace="a"
variation       7147
                /gene="IRAK4"
                /frequency="0.01"
                /replace="a"
variation       7234
                /gene="IRAK4"
                /frequency="0.04"
                /replace="c"
repeat_region   7367..8636
                /rpt_family="Mariner"
                /rpt_type=dispersed
variation       7392
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"
variation       7597
                /gene="IRAK4"
                /frequency="0.01"
                /replace="c"
variation       7794
                /gene="IRAK4"
                /frequency="0.09"
                /replace="g"
```

```
RESULT 15
AC093012

Query Match     51.6%;     Score 151.8;     DB 9;     Length 33033;
Best Local Similarity   95.7%;     Pred. No. 1.6e-28;
Matches 156;  Conservative 0;  Mismatches 7;  Indels 0;  Gaps 0;

Qy    123  TCAGTTTCACATAAGGAGGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTCTGA  182
           |||||||| |||||| |||||||||||||||||||||| ||||||||||||||||||||| ||
Db  14142  TTACTTACTTTTAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGA  14201

Qy    183  ATTACTGTTTGACTGGGGCGACCACCACAAAATTCCACACGTTGGTTGGTCGATCTTTTTGAT  242
           |||||||||||||||||||  |||||||||||| ||||| |||||  | |||||| ||||||||
Db  14202  ATTACTGTTTGACTGGGGCCACCACAAAATTGCACAGTTGGTGTGATCTTGTGGGATCTTTTTGAT  14261

Qy    243  CCAAAATGAATTTTTTTGCTCCCTGCGGAGTCTTTTTGCTCCCAGAT  285
           ||||||||||||||||||||||||||| |||||||||||||||| ||
Db  14262  CCAAAATGAATTTTTTTGCTCCCTGCGAGTCTTTTTGCTCCCCAGGT  14304
```

REFERENCE   2 (bases 1 to 118572)
  AUTHORS   Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
            Escotto,M., Falle,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
            Frantz,P., Gabisi,A., Gao,J., Garcia,W., Garner,T., Garza,N.,
            Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
            Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Ioshikhes,I., Jackson,L., Jacobson,B., Jia,Y.,
            Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
            King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
            Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
            Lieu,C., Liu,J., Liu,W., Louiseged,H., Lozado,R.J., Lu,X.,
            Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
            Marondel,I., Martin,J., Martindale,A., Martinez,E., Massey,E.,
            Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
            Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
            Mobabat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
            Neal,D., Nelson,D., Newton,J., Newton,Y., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,E., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Tang,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,J., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,A., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
            Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished

REFERENCE   3 (bases 1 to 118572)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

REFERENCE   4 (bases 1 to 118572)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

REFERENCE   5 (bases 1 to 118572)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

REFERENCE   6 (bases 1 to 118572)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

REFERENCE   7 (bases 1 to 118572)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

REFERENCE   8 (bases 1 to 118572)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One

COMMENT   Baylor Plaza, Houston, TX 77030, USA
            On Jun 29, 2002 this sequence version replaced gi:21535906.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
            ml.

FEATURES   Location/Qualifiers
     source   1..118572
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-210N13"
     misc_feature   complement(1..2004)
                     /note="overlaps bases 1..2004 of clone AC016143"
                     /function="clone overlap"
     repeat_region   869..903
                     /rpt_family="HSMAR1"
     repeat_region   1047..1467
                     /rpt_family="MLT1K"
     repeat_region   1644..1760
                     /rpt_family="MIR"
     repeat_region   1985..2295
                     /rpt_family="MIR"
     repeat_region   2296..2325
                     /rpt_family="AluY"
     repeat_region   complement(2730..3024)
                     /rpt_family="AT_rich"
     repeat_region   4092..4211
                     /rpt_family="Alusg"
     repeat_region   4389..4415
                     /rpt_family="MER5A"
     repeat_region   4502..4561
                     /rpt_family="AT_rich"
     repeat_region   complement(4612..4778)
                     /rpt_family="MIR"
     repeat_region   5859..5926
                     /rpt_family="AT_rich"
     repeat_region   complement(6385..6636)
                     /rpt_family="Alusg"

```
repeat_region    complement(7252..7562)
                 /rpt_family="L2"
repeat_region    7574..8066
                 /rpt_family="MLT1F"
repeat_region    complement(8266..8438)
                 /rpt_family="MIR"
repeat_region    8885..8927
                 /rpt_family="AT_rich"
repeat_region    10113..10252
                 /rpt_family="L1M4"
repeat_region    10433..10726
                 /rpt_family="AluY"
repeat_region    11181..11552
                 /rpt_family="THE1C"
repeat_region    complement(11739..11889)
                 /rpt_family="MER5B"
repeat_region    11910..12124
                 /rpt_family="MER58A"
repeat_region    12249..12617
                 /rpt_family="THE1C"
repeat_region    14290..14310
                 /rpt_family="AT_rich"
repeat_region    14363..14670
                 /rpt_family="AluSq"
repeat_region    complement(14742..15039)
                 /rpt_family="AluSx"
repeat_region    15669..15691
                 /rpt_family="(TTG)n"
repeat_region    complement(15692..15972)
                 /rpt_family="AluSc"
repeat_region    16265..16352
                 /rpt_family="L1PA16"
repeat_region    16353..16631
                 /rpt_family="L1PA16"
repeat_region    16632..17007
                 /rpt_family="AluJo"
repeat_region    17329..17356
                 /rpt_family="L1PA16"
```

```
Query Match      51.6%;    Score 151.8;  DB 9;   Length 118572;
Best Local Similarity    95.7%;    Pred. No. 1.4e-28;
Matches 156;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy    123 TCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGA 182
          | |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38156 TTACTTACTTTTAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGA 38215

Qy    183 ATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGTGAATCTTGGTGGATCTTTTGAT 242
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  38216 ATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGTGAATCTTGGTGGATCTTTTGAT 38275

Qy    243 CCAAAATGAATTTTTTGCTCCTGCGGAGTCTTTTGCTCCCAGAT 285
          ||||||||||||||||||||||||||||||||||||||||||||
Db  38276 CCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCCAGGT 38318
```

```
OM nucleic - nucleic search, using sw model

                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:             January 17, 2004, 23:50:52 ; Search time 51 Seconds
                                                 (without alignments)
                                                 2544.444 Million cell updates/sec

Title:              US-10-001-254-5
Perfect score:      294
Sequence:           1 acatatgtgcgctgcccaa..........tgctccaagatgctgtccc 294

Scoring table:      IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0
Searched:           569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:    1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :          Issued_Patents NA:*
                    1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                    2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                    3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                    4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                    5: /cgn2_6/ptodata/1/ina/PcTUS_COMB.seq:*
                    6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                    SUMMARIES
```

Result No.	Score	% Query Match	Length	Match DB ID	Description
1	100.0	100.0	833	4	US-09-166-350-10	Sequence 10, Appl
2	40.4	13.7	1879	4	US-09-601-091-1	Sequence 1, Appli
3	40.4	13.7	2024	4	US-09-601-091-3	Sequence 3, Appli
4	40.4	13.7	2024	4	US-09-398-395A-51	Sequence 51, Appl
5	40.4	13.7	2024	4	US-09-887-586A-51	Sequence 51, Appl
6	40.4	13.7	2024	4	US-09-895-752-51	Sequence 51, Appl
7	40.4	13.7	2024	4	US-09-903-012B-51	Sequence 51, Appl
8	33.8	11.5	580073	4	US-08-545-528D-1	Sequence 1, Appli
9	33.6	11.4	2393	4	US-08-961-527-90	Sequence 90, Appl
10	32.8	11.2	202001	4	US-09-734-674-3	Sequence 3, Appli
11	32.6	11.1	1728	4	US-09-096-731A-1	Sequence 1, Appli
12	32.4	11.0	98844	4	US-09-791-211-10	Sequence 10, Appl
13	32.2	11.0	4402	4	US-09-484-970B-135	Sequence 135, App
14	32	10.9	2274	4	US-09-107-532A-3535	Sequence 3535, Ap
15	31.8	10.8	1557	4	US-09-134-001C-1614	Sequence 1614, Ap
16	31.8	10.8	2507	4	US-09-620-312D-120	Sequence 120, App
17	31.8	10.8	1664976	4	US-08-916-421B-1	Sequence 1, Appli
18	31.4	10.7	12720	2	US-08-403-866-11	Sequence 11, Appl
19	31	10.5	2126	2	US-08-545-745-1	Sequence 1, Appli
20	30.8	10.5	1550	4	US-09-427-501-1	Sequence 1, Appli
21	30.8	10.5	15016	4	US-09-601-198-60	Sequence 60, Appl
22	30.8	10.5	319608	4	US-09-539-333D-1	Sequence 1, Appli
23	30.8	10.5	319608	4	US-09-679-409-1	Sequence 1, Appli
24	30.8	10.5	1230025	4	US-09-198-452A-1	Sequence 1, Appli
25	30.6	10.4	1479	4	US-09-328-352-897	Sequence 897, App
26	30.4	10.3	1362	3	US-08-929-329-4	Sequence 4, Appli
27	30.2	10.3	2254	2	US-08-635-066-1	Sequence 1, Appli
28	30	10.2	195	1	US-08-324-243-20	Sequence 20, Appl
29	30	10.2	195	1	US-08-532-390-20	Sequence 20, Appl
30	30	10.2	195	3	US-08-717-294-20	Sequence 20, Appl
31	30	10.2	195	5	PCT-US95-11511-20	Sequence 20, Appl
32	30	10.2	1632	1	US-08-324-243-34	Sequence 34, Appl
33	30	10.2	1632	1	US-08-532-390-34	Sequence 34, Appl
34	30	10.2	1632	3	US-08-717-294-34	Sequence 34, Appl
35	30	10.2	1632	5	PCT-US95-11511-34	Sequence 34, Appl
36	30	10.2	2481	1	US-08-324-243-35	Sequence 35, Appl
37	30	10.2	2481	1	US-08-532-390-35	Sequence 35, Appl
38	30	10.2	2481	3	US-08-717-294-35	Sequence 35, Appl
39	30	10.2	2481	5	PCT-US95-11511-35	Sequence 35, App
40	30	10.2	2571	5	US-09-552-950-4	Sequence 4, Appli
41	30	10.2	11770	4	US-08-961-527-172	Sequence 172, App
42	29.8	10.1	4648	4	US-09-207-914-24	Sequence 24, Appl
43	29.8	10.1	4655	4	US-09-207-914-20	Sequence 20, Appl
44	29.6	10.1	1356	1	US-08-592-936B-22	Sequence 22, Appl
45	29.6	10.1	1356	1	US-08-788-928A-2	Sequence 2, Appli

ALIGNMENTS

DNA of Sea ID#5

```
RESULT 1
US-09-166-350-10
; Sequence 10, Application US/09166350A
; Patent No. 6440663
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Chen, Yao
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alex
; TITLE OF INVENTION: Renal Cancer Associated Antigens and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: L0461/7051
; CURRENT APPLICATION NUMBER: US/09/166,350A
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: US 09/166,350
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 833
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-166-350-10

 Query Match        100.0%;  Score 294;  DB 4;  Length 833;
 Best Local Similarity  100.0%;  Pred. No. 3.4e-82;
 Matches 294;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy      1 ACATATGTGCGCTGCCCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     74 ACATATGTGCGCTGCCCAATGTTGGACTAATTAGGAAGCTGTCAGATTTTATTGATCCT 133

Qy     61 CAAGAAGGAGATGGAAGAGAAGTTAGCTGTTAGCCTATTAAAAAAACCATCTGGTGATGATGATAGAGATAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134 CAAGAAGGAGATGGAAGAGAAGTTAGCTGTTAGCCTATTAAAAAAACCATCTGGTGATGATGATAGAGATAC 193

Qy    121 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAGTCCCACTTCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194 AATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAGTCCCACTTCT 253

Qy    181 GAATTACTGTTTGACTGGGGGCACCAACAAATTGCACAGTTCGTGACTCTTTGTGTGGATCTTTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 GAATTACTGTTTGACTGGGGGCACCAACAAATTGCACAGTTCGTGACTCTTTGTGTGGATCTTTG 313

Qy    241 ATCCAAAAATGAATTTTTTTTGCTCCTGCGAGTCTTTTGCTTCCCAGAGATGCTGTTCCC 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 ATCCAAAAATGAATTTTTTTTGCTCCTGCGAGTCTTTTGCTTCCCAGAGATGCTGTTCCC 367
```

RESULT 2

US-09-601-091-1

; Sequence 1, Application US/09601091
; Patent No. 6342380
; GENERAL INFORMATION:
;　APPLICANT: Colby, S. M. et al.
;　TITLE OF INVENTION: Germacrene C Synthase Gene of Lycopersicon Esulentum
;　FILE REFERENCE: 4630-55758
;　CURRENT APPLICATION NUMBER: US/09/601,091
;　CURRENT FILING DATE: 2000-09-19
;　PRIOR APPLICATION NUMBER: PCT/US99/02133
;　PRIOR FILING DATE: 1999-02-02
;　PRIOR APPLICATION NUMBER: US 60/073,579
;　PRIOR FILING DATE: 1998-02-02
;　NUMBER OF SEQ ID NOS: 6
;　SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 1
;　LENGTH: 1879
;　TYPE: DNA
;　ORGANISM: Lycopersicon esculentum
;　FEATURE:
;　NAME/KEY: CDS
;　LOCATION: (39)..(1685)

US-09-601-091-1

Query Match　　　　　13.7%;　Score 40.4;　DB 4;　Length 1879;
Best Local Similarity　50.5%;　Pred. No. 0.0042;
Matches　98;　Conservative　0;　Mismatches 96;　Indels　0;　Gaps　0;

Qy　　45 AGATTTTATTGATCCTCAAGAAGGAGGATGGAAGAAGAGTTAGCTGTAGCTATTAAAAAACCATC 104
　　　　 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　761 AGAGCTTAGTGATCTTACAAGGTGGTGGAAAGAGATTTGGATTTGCAAATAAATATCCATA 820

Qy　　105 TGGTGATGATAGATACAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGG 164
　　　　 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　821 TGCAAGAGACAGGTTGGTTGAGTGTTACTTCTGGATATTAGGGAGTGTATTTTGAGCCAAA 880

Qy　　165 AAAAAGTCCCACTTCTGAATTACTCGTTTGACTGGGGCGCACCACAAAATTGCACAGTTGGTGA 224
　　　　 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　881 ATATAGTCGTGCGAGAAAAATGATGACAAAAGTACTCAACCTGACCTCCATTATTGACGA 940

Qy　　225 TCTTGTGGGATCTTT 238
　　　　 |||||||||||||
Db　　941 CACTTTTGATGCTT 954

RESULT 3

US-09-601-091-3

; Sequence 3, Application US/09601091
; Patent No. 6342380
; GENERAL INFORMATION:
;　APPLICANT: Colby, S. M. et al.
;　TITLE OF INVENTION: Germacrene C Synthase Gene of Lycopersicon Esulentum
;　FILE REFERENCE: 4630-55758
;　CURRENT APPLICATION NUMBER: US/09/601,091
;　CURRENT FILING DATE: 2000-09-19
;　PRIOR APPLICATION NUMBER: PCT/US99/02133
;　PRIOR FILING DATE: 1999-02-02
;　PRIOR APPLICATION NUMBER: US 60/073,579
;　PRIOR FILING DATE: 1998-02-02
;　NUMBER OF SEQ ID NOS: 6
;　SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 3
;　LENGTH: 2024
;　TYPE: DNA
;　ORGANISM: Lycopersicon esculentum
;　FEATURE:
;　NAME/KEY: CDS
;　LOCATION: (32)..(1678)

US-09-601-091-3

Query Match　　　　　13.7%;　Score 40.4;　DB 4;　Length 2024;

RESULT 4

US-09-398-395A-51

; Sequence 51, Application US/09398395A
; Patent No. 6468772
; GENERAL INFORMATION:
;　APPLICANT: Chappell, Joseph
;　APPLICANT: No. 6468772I, Joseph P.
;　APPLICANT: Starks, Courtney M.
;　APPLICANT: Manna, Kathleen R.
;　TITLE OF INVENTION: SYNTHASES
;　FILE REFERENCE: 07678-025001
;　CURRENT APPLICATION NUMBER: US/09/398,395A
;　CURRENT FILING DATE: 1999-09-17
;　PRIOR APPLICATION NUMBER: 60/100,993
;　PRIOR FILING DATE: 1998-09-18
;　PRIOR APPLICATION NUMBER: 60/130,628
;　PRIOR FILING DATE: 1999-04-22
;　PRIOR APPLICATION NUMBER: 60/150,262
;　PRIOR FILING DATE: 1999-08-23
;　NUMBER OF SEQ ID NOS: 58
;　SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 51
;　LENGTH: 2024
;　TYPE: DNA
;　ORGANISM: Lycopersicon esculentum
;　FEATURE:
;　NAME/KEY: CDS
;　LOCATION: (32)..(1675)
;　OTHER INFORMATION: VFNT germacrene C synthase

US-09-398-395A-51

Query Match　　　　　13.7%;　Score 40.4;　DB 4;　Length 2024;
Best Local Similarity　50.5%;　Pred. No. 0.0043;
Matches　98;　Conservative　0;　Mismatches 96;　Indels　0;　Gaps　0;

Qy　　45 AGATTTTATTGATCCTCAAGAAGGAGGATGGAAGAAGAGTTAGCTGTAGCTATTAAAAAACCATC 104
　　　　 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　754 AGAGCTTAGTGATCTTACAAGGTGGTGGAAAGAGATTTGGATTTGCAAATAAATATCCATA 813

Qy　　105 TGGTGATGATAGATACAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGG 164
　　　　 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　814 TGCAAGAGACAGGTTGGTTGAGTGTTACTTCTGGATATTAGGGAGTGTATTTTGAGCCAAA 873

Qy　　165 AAAAAGTCCCACTTCTGAATTACTCGTTTGACTGGGGCGCACCACAAAATTGCACAGTTGGTGA 224
　　　　 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　874 ATATAGTCGTGCGAGAAAAATGATGACAAAAGTACTCAACCTGACCTCCATTATTGACGA 933

Qy　　225 TCTTGTGGGATCTTT 238
　　　　 |||||||||||||
Db　　934 CACTTTTGATGCTT 947

Best Local Similarity　50.5%;　Pred. No. 0.0043;
Matches　98;　Conservative　0;　Mismatches 96;　Indels　0;　Gaps　0;

Qy　　45 AGATTTTATTGATCCTCAAGAAGGAGGATGGAAGAAGAGTTAGCTGTAGCTATTAAAAAACCATC 104
　　　　 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　754 AGAGCTTAGTGATCTTACAAGGTGGTGGAAAGAGATTTGGATTTGCAAATAAATATCCATA 813

Qy　　105 TGGTGATGATAGATACAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGG 164
　　　　 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　814 TGCAAGAGACAGGTTGGTTGAGTGTTACTTCTGGATATTAGGGAGTGTATTTTGAGCCAAA 873

Qy　　165 AAAAAGTCCCACTTCTGAATTACTCGTTTGACTGGGGCGCACCACAAAATTGCACAGTTGGTGA 224
　　　　 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　874 ATATAGTCGTGCGAGAAAAATGATGACAAAAGTACTCAACCTGACCTCCATTATTGACGA 933

Qy　　225 TCTTGTGGGATCTTT 238
　　　　 |||||||||||||
Db　　934 CACTTTTGATGCTT 947

RESULT 5

```
US-09-887-586A-51
  Sequence 51, Application US/0987586A
  Patent No. 6495354
  CURRENT FILING DATE: 2001-06-22
  PRIOR APPLICATION NUMBER: 09/398,395
  PRIOR FILING DATE: 1999-09-17
  PRIOR APPLICATION NUMBER: 60/130,628
  PRIOR FILING DATE: 1999-04-22
  PRIOR APPLICATION NUMBER: 60/150,262
  PRIOR FILING DATE: 1999-08-23
  NUMBER OF SEQ ID NOS: 58
  SOFTWARE: FastSEQ for Windows Version 3.0
  SEQ ID NO 51
    LENGTH: 2024
    TYPE: DNA
    ORGANISM: Lycopersicon esculentum
    FEATURE:
    NAME/KEY: CDS
    LOCATION: (32)...(1675)
  OTHER INFORMATION: VFNT germacrene C synthase
US-09-887-586A-51

Query Match             13.7%;  Score 40.4;  DB 4;  Length 2024;
Best Local Similarity   50.5%;  Pred. No. 0.0043;
  Matches   98;  Conservative  0;  Mismatches  96;  Indels   0;  Gaps   0;

Qy        45 AGATTTATTGATCCTCAAGAAGGATGGAAGAGAAGTTAGCTGTAGCTATTAAAAAACCATC 104
             ||| || ||||    |||  ||   |    ||| ||||| |||    ||||||||||||||
Db       754 AGAGCTTAGTGATCTTACACAAGGTGGTGGGAAAGAGATTTGGAATTTTGCAAATAATCCATA 813

Qy       105 TGGTGATGATAGATACAAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGG 164
             |||| |||||||||    |||  || |||||| ||||| ||| ||| |||| || |||||||
Db       814 TGCAAGAGACAGGTTCGGTTGAGTGTTACTTCTGGATTATTAGGAGTGTATATTTGAGCCAAA 873

Qy       165 AAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGGTGA 224
             || |||||| |||||||| | ||||| ||| |||||  |||||||| |||| |||||||||
Db       874 ATATAGTCGTCCGAGAGAAAAATGATGACAAAAAGTACTCAACCTGACCTCCATTATTGACGA 933

Qy       225 TCTTGTGGATCTTT 238
             ||| |||||| |||
Db       934 CACTTTTGATGCTT 947
```

RESULT 6
```
US-09-895-752-51
  Sequence 51, Application US/0995752
  Patent No. 6559297
  GENERAL INFORMATION:
  APPLICANT: Chappell, Joseph
  APPLICANT: No. 6559297
  APPLICANT: Starks, Courtney M.
  APPLICANT: Manna, Kathleen R.
  TITLE OF INVENTION: SYNTHASES
  FILE REFERENCE: 07678-025001
  CURRENT APPLICATION NUMBER: US/09/895,752
  CURRENT FILING DATE: 2001-06-29
  PRIOR APPLICATION NUMBER: 09/398,395
  PRIOR FILING DATE: 1999-09-17
  PRIOR APPLICATION NUMBER: 60/100,993
  PRIOR FILING DATE: 1998-09-18
  PRIOR APPLICATION NUMBER: 60/130,628
  PRIOR FILING DATE: 1999-04-22
  PRIOR APPLICATION NUMBER: 60/150,262
  PRIOR FILING DATE: 1999-08-23
  NUMBER OF SEQ ID NOS: 58
```

```
  SOFTWARE: FastSEQ for Windows Version 3.0
  SEQ ID NO 51
    LENGTH: 2024
    TYPE: DNA
    ORGANISM: Lycopersicon esculentum
    FEATURE:
    NAME/KEY: CDS
    LOCATION: (32)...(1675)
  OTHER INFORMATION: VFNT germacrene C synthase
US-09-895-752-51

Query Match             13.7%;  Score 40.4;  DB 4;  Length 2024;
Best Local Similarity   50.5%;  Pred. No. 0.0043;
  Matches   98;  Conservative  0;  Mismatches  96;  Indels   0;  Gaps   0;

Qy        45 AGATTTATTGATCCTCAAGAAGGATGGAAGAGAAGTTAGCTGTAGCTATTAAAAAACCATC 104
             ||| || ||||    |||  ||   |    ||| ||||| |||    ||||||||||||||
Db       754 AGAGCTTAGTGATCTTACACAAGGTGGTGGGAAAGAGATTTGGAATTTTGCAAATAATCCATA 813

Qy       105 TGGTGATGATAGATACAAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGG 164
             |||| |||||||||    |||  || |||||| ||||| ||| ||| |||| || |||||||
Db       814 TGCAAGAGACAGGTTCGGTTGAGTGTTACTTCTGGATTATTAGGAGTGTATATTTGAGCCAAA 873

Qy       165 AAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCACCACAAAATTGCACAGTTGGGTGA 224
             || |||||| |||||||| | ||||| ||| |||||  |||||||| |||| |||||||||
Db       874 ATATAGTCGTCCGAGAGAAAAATGATGACAAAAAGTACTCAACCTGACCTCCATTATTGACGA 933

Qy       225 TCTTGTGGATCTTT 238
             ||| |||||| |||
Db       934 CACTTTTGATGCTT 947
```

RESULT 7
```
US-09-903-012B-51
  Sequence 51, Application US/0903012B
  Patent No. 6559656
  GENERAL INFORMATION:
  APPLICANT: Chappell, Joseph
  APPLICANT: No. 6559656
  APPLICANT: Starks, Courtney M.
  APPLICANT: Manna, Kathleen R.
  TITLE OF INVENTION: SYNTHASES
  FILE REFERENCE: 07678-025001
  CURRENT APPLICATION NUMBER: US/09/903,012B
  CURRENT FILING DATE: 2001-07-11
  PRIOR APPLICATION NUMBER: 60/100,993
  PRIOR FILING DATE: 1998-09-18
  PRIOR APPLICATION NUMBER: 60/130,628
  PRIOR FILING DATE: 1999-04-22
  PRIOR APPLICATION NUMBER: 60/150,262
  PRIOR FILING DATE: 1999-08-23
  NUMBER OF SEQ ID NOS: 58
  SOFTWARE: FastSEQ for Windows Version 3.0
  SEQ ID NO 51
    LENGTH: 2024
    TYPE: DNA
    ORGANISM: Lycopersicon esculentum
    FEATURE:
    NAME/KEY: CDS
    LOCATION: (32)...(1675)
  OTHER INFORMATION: VFNT germacrene C synthase
US-09-903-012B-51

Query Match             13.7%;  Score 40.4;  DB 4;  Length 2024;
Best Local Similarity   50.5%;  Pred. No. 0.0043;
  Matches   98;  Conservative  0;  Mismatches  96;  Indels   0;  Gaps   0;

Qy        45 AGATTTATTGATCCTCAAGAAGGATGGAAGAGAAGTTAGCTGTAGCTATTAAAAAACCATC 104
             ||| || ||||    |||  ||   |    ||| ||||| |||    ||||||||||||||
Db       754 AGAGCTTAGTGATCTTACACAAGGTGGTGGGAAAGAGATTTGGAATTTTGCAAATAATCCATA 813

Qy       105 TGGTGATGATAGATACAAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGG 164
```

Db   814 TGCAAGAGACAGGTTGGTTGGTTGAGTGTTACTTCTGGATTATTAGGAGTGTGTATTTTGAGCCAAA 873

Qy   165 AAAAAGTCCCACTTCTGAATTACTGTTTGGGGGCACCACAAAAATTGCACAGTTGGGTGA 224

Db   874 ATATAGTGTCGTGCGAGAAAAATGATGACAAAAAGTACTCAACCTGACCTCCATTATTGACGA 933

Qy   225 TCTTGTGGAACTTT 238

Db   934 CACTTTGATGCTT 947

RESULT 8
; US-08-545-528D-1/c
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin version 3.1
;
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match   11.5%; Score 33.8; DB 4; Length 580073;
Best Local Similarity   53.4%; Pred. No. 4.6;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy   143 TTGAAGCATTACTTCAAACTGGAAAAACCAAAACCTTGATTGAATCATTATTATTGACTGTTTGACTGGGGCA 202
           |||||||||||||||||||| |||||||| ||||||| ||||||| |||||||||| ||||
Db 431703 TTGAAGCAATTGTTGAAAACCAAAACCTTGATTCTTTTTGATCCAAAAATGAATCATTATTTTTTGCTC 431644

Qy   203 CCACAAAATTGCACAGTTGGGTGGTCGATCTTGTGTGGATCTTTTTGATCCAAAAATGAATTTTTTTGCTC 262
           ||| ||||| | ||||| | |||||||||||||| |||||
Db 431643 TTAATCGCTACTCAATTACTCCTACAGTGTGACACAAAAAAACTATTGTTAGAAG 431584

Qy   263 CTGCGAGTCTTTT 275
           ||| |||||
Db 431583 CTAACAGTCTTAT 431571

RESULT 9
; US-08-961-527-90/c
; Sequence 90, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 2393 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-90

Query Match   11.4%; Score 33.6; DB 4; Length 2393;
Best Local Similarity   48.9%; Pred. No. 0.6;
Matches 90; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy   37 AAGCTGTCAGATTTTTATTGATCCTCAAGAAGGATGGAAGGAAGTTTAGCTGTTAGCTATTAAA 96
           |||||||||||||| |||||||||||||||||||||||||||| || || |||||| |||||
Db   1113 AACCTTTTAGTAACCAACCGGAGAACTAAAAGACTCCGCAACAAAAAGACGGATATTCCTATTAAG 1054

Qy   97 AAACCATCTGGTGATGATAGATACACAATCAGTTTCACATAAGGAGGAGATTTGAAGCATTACTT 156
           || || || ||| ||| || || ||| || || || |||||| ||||| |||
Db   1053 ATAGGAACTAGTGATGTGGCAATTACTGATTTTGAAATAATGTAATGTTAGCAATGAT 994

Qy   157 CAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTGACTGGGGCCACCACAAAATTGCACA 216
           ||| |||| || || ||| ||||||||||| ||||||||||||| ||||||
Db   993 GTTACTTCTAAAGAAAATATCTGCTAAAATTGCAAAATAATTGAGAGACTAACAACTTGACA 934

Qy   217 GTTG 220
           ||||
Db   933 AATG 930

RESULT 10
; US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. 6498022
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

Query Match   11.2%; Score 32.8; DB 4; Length 202001;
Best Local Similarity   58.0%; Pred. No. 6.2;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy   62 AAGAAGGATGGAAGGAAGTTTAGCTGTAGCTATTAAAAAAACCATCTGGTGTGATGATAGATACA 121
           ||| ||| || | |||| |||| | ||| |||||||||||| |||||| |||| |||||||
Db 91622 AAGAGAGTAACAAAAAAATTAGCTCTGTACCTTATAAAATATACACTAAGGGGGAAAAGATATACC 91563

Qy   122 ATCAGTTTCACATAAGGAGAGATTTGAACCATTACTTCAAAC 161
             ||| |||| ||||| ||| || |||| || |||
Db   91562 TGCAATTTCTTCTAAAAGTGTTGGCATCCATGTCCAGCAAC 91523

RESULT 11
US-09-096-731A-1
; Sequence 1, Application US/09096731A
; Patent No. 6489454
; GENERAL INFORMATION:
; APPLICANT: Liu, Qiong
; APPLICANT: Hengartner, Michael O.
; APPLICANT: Bogaert, Thierry Andre Oliver Eddy
; APPLICANT: Van Criekinge, Wim Maria Rene
; TITLE OF INVENTION: ENGULFMENT GENE AND USES THEREOF
; FILE REFERENCE: CSHL97-07pA
; CURRENT APPLICATION NUMBER: US/09,096,731A
; CURRENT FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: US 60/072,324
; PRIOR FILING DATE: 1998-01-23
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)...(1500)
US-09-096-731A-1

Query Match   11.1%; Score 32.6; DB 4; Length 1728;
Best Local Similarity 51.7%; Pred. No. 1.1;
Matches 74; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy   46 GATTTTATTGATCCTCAAGAAGGATGCGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCT 105
             ||||| ||| || ||| | || | ||||| || || || ||||| ||||||
Db   1132 GATTCGTTTGATCCCAGAGCTGCGAGAAAAAAAGACGCACTGCAGCTGAGTATATAATCTC 1191

Qy   106 GGTGATGATAGATACAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGA 165
             |||| || ||||| |||||| |||||||||||||||||||||||||||||||||||||||
Db   1192 GGTGCGGACTTCCTCAGTGGCATTCAAAATGGTAAAGAAGCACCACCATCAGCCTCCGCT 1251

Qy   166 AAAAGTCCCACTTCTGAATTACT 188
             ||| | || | | ||||| |
Db   1252 GAACTTCTCCGTTCTGAAGCAAT 1274

RESULT 12
US-09-791-211-10/c
; Sequence 10, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
; FILE REFERENCE: RTS-0205
; CURRENT APPLICATION NUMBER: US/09/791,211
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
; LENGTH: 98844
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: unsure
; LOCATION: 24962
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 64383
; OTHER INFORMATION: unknown
; NAME/KEY: unsure

; LOCATION: 65468
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65470
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65469
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65471
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 87130
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 89049
; OTHER INFORMATION: unknown
; OTHER INFORMATION:
US-09-791-211-10

Query Match   11.0%; Score 32.4; DB 4; Length 98844;
Best Local Similarity 53.1%; Pred. No. 6.2;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy   51 TATTGACCTCAAGAAGGATGCGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGA 110
             | ||| | || |||| || |||| | |||| || || ||||| ||||||| || ||
Db   7415 TATATATTTTCATGTATAAATGGAAAAAAATCCGTACTTCGTTATTAATTGACATGCTATTT 7356

Qy   111 TGATAGATACAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAAG 170
             ||| || | |||||||| ||||| | | | |||||||| | ||| || |||||| ||
Db   7355 TAAATCATTAAATTTATTTCGGTAATGAGAGATTTTAATCAAATATTTTTGAAATAGGGAAA 7296

Qy   171 TCCCACTTCT 180
             |||||| | |
Db   7295 ACATACATAT 7286

RESULT 13
US-09-484-970B-135
; Sequence 135, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 135
; LENGTH: 4402
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186 252875 (2874502CB1)
US-09-484-970B-135

Query Match   11.0%; Score 32.2; DB 4; Length 4402;
Best Local Similarity 51.9%; Pred. No. 2.1;
Matches 97; Conservative 0; Mismatches 88; Indels 2; Gaps 1;

Qy   51 TATTGACCTCAAGAAGGATGCGAAGAAGTTAGCTGTAGCTATTAAAAAACCATCTGGTGA 110
             ||| | || | |||| ||| | | |||| | ||| || ||||||||| || |||
Db   2771 ATAACATATGAAAATTCAGTAGAATAGTGTGTTTCATAGGAGCTTTTAACTGGTAAA 2830

Qy   110 ATGATAGATACAATCAGTTTCACATAAGGAGAGATTTGAAGCATTACTTCAAACTGGAAAAA 169
             |||||||||| ||||||||||||||||||||||||||||||| |||| |||||||| | ||
Db   2771 ATAACATATGAAAATTCAGTAGAATAGTGTGTTTCATAGGAGCTTTTAACTGGTAAA 2830

Qy   170 GTCCCACTT--CTGAATTACTGTTGACTGGGGCACCACAAATTGCACAGTTGGTGATCT 227
             | ||| | | | || |||| | || || ||| | ||| || | || | |
Db   2831 ATATTACTTGCACGAGAGTACTTGATGTATGGTTATCCTGAAATTTCGGAGTATTTGGTGT 2890

Qy   228 TGTGGATCTTTTGATCCAAAATGAATTTTTTGCTCCTGCGAGTCTTTTGCTCCCAGATGC 287

```
Db   2891 GTTCTTTTGTCTAAAAATAGTCTGTTTTTGTCAGTTCCTTCAGAATATATTTATTCTGAAGA 2950
              ||| || ||| |||| |||| |  ||| |||  ||||  ||| |||||   || ||||
Qy    288 TGTTCCC 294
              |||||
Db   2951 TTGTCCC 2957
```

RESULT 14
US-09-107-532A-3535
; Sequence 3535, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;   ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;   STREET: 100 Beaver Street
;   CITY: Waltham
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02354
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: CD/ROM ISO9660
;   COMPUTER: PC
;   OPERATING SYSTEM: <Unknown>
;   SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/107,532A
;   FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/085,598
;   FILING DATE: 14 May 1998
;   APPLICATION NUMBER: 60/051571
;   FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ariniello, Pamela Deneke
;   REGISTRATION NUMBER: 40,489
;   REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (781)893-5007
;   TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3535:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2274 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: circular
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   ORGANISM: Enterococcus faecium
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (B) LOCATION 1...2274
US-09-107-532A-3535
;   SEQUENCE DESCRIPION: SEQ ID NO: 3535:

Query Match   10.9%;   Score 32;   DB 4;   Length 2274;
Best Local Similarity   52.2%;   Pred. No. 1.9;
Matches 71; Conservative   0; Mismatches 65; Indels   0; Gaps   0;

Qy   30 AATTAGGAAGCTGTCAGATTTTATTTGATCCTCAGAAGGATGGAGAAGTTAGCTGTAGC 89
            |||| || | ||| ||| | ||| ||| || || ||||| | ||| ||||| ||
Db 1559 AATCACTTAGCCGCAAATTTTAGAGTTACTGGATAATGAACAAACAGAGACGTTACCAA 1618

Qy   90 TATTAAAAAACCATCTTGGTGATGATAGATACAAATCAGTTTCACATAAGGAGATTTGAAGC 149
            ||| | ||| ||| |||| | ||| || | || |||||||| || ||| | | |||||
Db 1619 AAGGAAAATAGGAAGGTCAGCGCGGTTCAAGCAAAGTAGCCATATGGTGTGATTTCGAGAA 1678
```

```
Qy    150 ATTACTTCAAAACTGGA 165
              ||| | |   ||| |
Db   1679 ATTTTGACAAAAAGAA 1694
```

RESULT 15
US-09-134-001C-1614
; Sequence 1614, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1614
;   LENGTH: 1557
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1614

Query Match   10.8%;   Score 31.8;   DB 4;   Length 1557;
Best Local Similarity   48.6%;   Pred. No. 1.9;
Matches 87; Conservative   0; Mismatches 92; Indels   0; Gaps   0;

Qy   87 AGCTATTAAAAAACCATCTGGTGATGATAGATACAAATCAGTTTCACATAAGGAGATTTGA 146
            ||| || || | | || ||| | ||| ||| || |||| | || ||| | |||
Db   1092 AGTAATTAATCAAAACAATTTGATAAAAATCAAAAAATTATATTGAAATTGGTAAAAAGA 1151

Qy   147 AGCATTACTTCAAACTGGAAAAAGTCCCACTTCTGAATTACTGTTTTGACTGGGGCACCAC 206
              |||||| || || | || || || | || ||||| || ||| |||| | | ||| ||
Db   1152 AGGCAAACTAGAGACTGGGTCGGTGGGTGAACAGAGATTCTACCGGTTATTCATTGAACAAC 1211

Qy   207 AAATTGCACAGTTGGTGATCTTGTGGATCTTTTGATCCAAAATGAATTTTTTGCTCCTG 265
             |||| | || | | ||| ||||| || |||||| || | | ||| || ||||| |||
Db   1212 GATTTTCTCCGGACTACAATCTGCGGATCGTATCATGCAAGAAGAAATTTTTGGACCAG 1270
```

Search completed: January 18, 2004, 00:40:18
Job time : 55 secs